Essays on the Economics of Deception and Debt

Dissertation
submitted to the
Faculty of Business, Economics and Informatics
of the University of Zurich

to obtain the degree of
Doktor der Wirtschaftswissenschaften, Dr. oec.
(corresponds to Doctor of Philosophy, PhD)

presented by

Andrin T. Bögli
from Seeberg, BE

approved in October 2017 at the request of

Prof. Dr. Alexander F. Wagner
Prof. Dr. Rajna Gibson Brandon

The Faculty of Business, Economics and Informatics of the University of Zurich hereby authorizes the printing of this dissertation, without indicating an opinion of the views expressed in the work.

Zurich, 25.10.2017

The Chairman of the Doctoral Board: Prof. Dr. Steven Ongena

Acknowledgments

I would like to express my deep gratitude to my supervisor Alexander F. Wagner for his guidance and generous support. His continuous feedback was of great value to this thesis. Moreover, I benefited a lot from his inspiring research enthusiasm and his ideas as a co-author.

I am very grateful to Joel Sobel for the outstanding hospitality during my time at the University of California, San Diego. The numerous enriching discussions improved my own thinking. In addition, I could learn greatly from our co-authorship. Many thanks also to Rajna Gibson Brandon for serving on my committee.

I am thankful to Richard J. Zeckhauser for inviting me to visit Harvard University and for his valuable remarks during that time. I have also profited from various discussions with Michel Habib, Steven Ongena, and Per Östberg.

Special thanks go to Felix Fattinger. It was a pleasure to learn from each other and to tackle the challenges of the program as well as of our joint paper together. I thank Ivan Petzev for sharing so many magnificent experiences.

I am grateful to my colleagues at the chair Sascha Behnk and Christoph Wenk Bernasconi for contributing to a relaxed but constructive atmosphere and for their helpfulness at all times. In addition, I thank fellow PhD students Fulvia Fringuellotti, René Hegglin, Lilia Mukhlynina, Diego Ostinelli, Stefano Ramelli, Thomas Richter, Cornelia Rösler, and Jiri Woschitz for valuable conversations.

Most importantly, I express my deepest gratitude to my family for their endless love and support.

Andrin T. Bögli, Zurich, August 2017

Contents

List of Tables

List of Figures

Part I

Dissertation Overview

Dissertation Overview

Deception is of major concern to society as illustrated by numerous cases of fraud in companies (e.g., by car manufacturers or banks), politics, sports, and many other areas. Deception has also drawn much attention in the philosophical, psychological, and economic literature. It has already been addressed by Homer and Plato, and later by David Hume, Adam Smith, Immanuel Kant, and Sigmund Freud, among others. This dissertation aims at discovering the mechanisms which determine individuals' proneness to unethical behavior. Understanding these mechanisms might ultimately help to build and maintain institutions that prevent misconduct.

This dissertation explores the economics of deception from three perspectives: micro, macro, and applied to sovereign debt. First, the micro analysis investigates a mental tool that people use to restructure their cognition when behaving unethically. Specifically, I study how people engage in *self-deception* in order to legitimate their deceitful behavior. Second, the macro approach addresses the *selection of economic models* from experimental data in the context of lying. It examines which preferences explain lying behavior well and how these preferences are distributed within the population. Finally, the applied approach focuses on advanced governments' incentives to borrow or deleverage in times of high sovereign indebtedness and analyzes the welfare implications of two *state-contingent sovereign debt* instruments. Each of the three independent research papers in this dissertation covers one perspective.

My first paper *When Self-Deception Promotes Unethical Advice: Biased Minds in Investment Consulting* studies self-deception as a self-serving device to promote unethical behavior in the financial advisory industry. Fraudulent advisor behavior is widespread and costly: 7.3% of the roughly 1.2 million financial advisors registered in the US between 2005 and 2015 have a record of misconduct whereas the median settlement paid in cases of misconduct amounts to \$40′000 (Egan et al. (2017)). In a first experiment on Amazon Mechanical Turk, I analyze whether investment advisors engage in self-deception by distorting their first-order beliefs about the likelihood that their behavior imposes harmful consequences on their clients.

The experimental methodology of this paper allows to decompose the effect of biased beliefs into two separate components: (i) a fundamental (incentive-independent) bias caused by an advisor's inability to assess the client's actual behavior and (ii) a self-deceptive bias due to motivated beliefs. I find advisors' self-deceptive biases to be significantly increasing in the magnitude of private benefits associated with a recommendation of an unfavorable investment opportunity. Once incentives to behave unethically are substantial, self-deception promotes dishonesty and leads to monetary gains for advisors at the cost of their clients. By contrast, the fundamental bias is constant and implies monetary losses for the advisor regardless of the direction of the bias.

A second experiment in a neutral strategic information transmission setting confirms the findings and suggests that people's engagement in self-deception has broad relevance for unethical behavior. Thanks to the general structure of the experiments, this paper's findings may also transfer to many other areas where asymmetric information leads to unethical behavior and people could engage in self-deception such as physicians proposing a suboptimal treatment, car mechanics recommending unnecessary repairs, and many more. Finally, the paper suggests that in an environment where the fundamental bias is comparably small and the incentives to lie are substantial, simply providing descriptive behavioral information to the informed party is enough to cost-efficiently foster ethical behavior.

My second paper *Model Selection from Experimental Data: Evidence from Individual Lying Behavior* is joint work with Joel Sobel and Alexander Wagner. We study how researchers should select a model of preferences from experimental data in the specific context of lying behavior. We show that offering a new model that organizes a particular data set better than a given candidate model does not provide evidence that the new model is useful in other situations. While the basic notion of a difference between in-sample fit and out-of-sample prediction is well-known, its relevance for experimental research may be underappreciated.

Using data from two different communication experiments, we run an empirical horse-race between five broad behavioral models used to explain lying behavior in the literature. Understanding the extent to which a narrow definition of utility fails to describe behavior is a necessary first step to understanding how to build and maintain institutions that lead to desirable outcomes even in the presence of temptations to lie. From a subset of the data, we estimate model parameters and assign participants to types. Then, we analyze how models that allow for multiple types fit the behavioral data in-sample and out-of-sample.

We find that models that explain behavior well in-sample do not necessarily perform well out-of-sample and vice versa. A model based on lying costs fits the data very well in-sample if there are only few observations, but does poorly even in-sample once more data are collected, and does very poorly compared to other models in predicting lying behavior out-of-sample. Combining in-sample and out-of-sample performance, a model consisting of three types (concerns for descriptive social norms, lying costs, and reference-dependent preferences) explains behavior best. We provide evidence that the share of each type within the whole population is quite stable across the two situations we study. Overall, our paper raises a caveat: Experimental research that bases results on an in-sample analysis with possibly few decisions per individual might produce misleading conclusions.

My third paper *Indebtedness, Interests, and Incentives: State-contingent Sovereign Debt Revisited* is joint work with Felix Fattinger. This paper studies state-contingent debt as an alternative refinancing instrument for advanced economies. In times of high sovereign indebtedness, increasing yields impose eminent debt roll-over risks. We analyze the welfare implications of two state-contingent debt instruments: puttable and GDP-to-debt-indexed bonds, both temporary in nature and intended to improve deleveraging feasibility.

In return for an insurance premium, puttable bonds offer protection against sovereign default, thereby internalizing the implicit risk-sharing mechanism inherited by the ECB's 'Outright Monetary Transactions' program. Similar to GDP-linked debt, bonds indexed to a country's GDP-to-debt ratio, henceforth 'GDR bonds', allow for consumption smoothing via state-contingent interest payments. In contrast to GDP-linked debt, GDR bonds permit competitive risk-return profiles even in the face of pessimistic growth outlooks.

We find that, in the presence of default costs, state-contingent bonds allow for substantial welfare improvements relative to standard sovereign debt. For risk-averse consumers, the counter-cyclical fiscal leeway created by GDR bonds dominates the interest savings provided by puttable bonds. We verify this preference order by calibrating our model to the five Eurozone countries most heavily affected by the debt crisis: Portugal, Ireland, Italy, Greece, and Spain. Finally, we discuss implied deleveraging incentives, limited commitment, and practical implementation issues for GDR bonds.

The structure of the dissertation is as follows: My three research papers are found in Part II. Part III contains the bibliography and Part IV presents my curriculum vitae.

Part II

Research Papers

1 When Self-deception Promotes Unethical Advice: Biased Minds in Investment Consulting*

"There is nothing worse than self-deception - when the deceiver is at home and always with you" - Plato, Cratylus 428d

1.1 Introduction

Ideas of self-deception in various forms have already been studied by Homer and Plato, and later by François de La Rochefoucauld, David Hume, Adam Smith, and Vilfredo Pareto. When deciding about committing unethical behavior such as fraud, corruption, deceptive communication or others, people trade off the monetary benefits that can be achieved against their inherent moral costs imposed by unethical behavior. Moral costs lead some people to act honestly even though this does not maximize their pecuniary payoff. Consequently, it could be monetarily beneficial to have lower moral costs. As Freud (1933) mentions, one way to relax the trade-off of monetary gains versus moral costs is that people subconsciously restructure cognition in order to reduce negative emotions.

Consider the situation of investment advisors whose salary is linked to the sale of some inferior financial products. By convincing themselves that clients are not going to follow their investment recommendations they might reduce negative emotions in order to legitimate unethical behavior. Egan et al. (2017) provide evidence that fraudulent advisor behavior is widespread and costly: 7.3% of the roughly 1.2 million financial advisors registered in the US between 2005 and 2015 have a record of misconduct whereas 29.3% of the misconducts are caused by a misrepresentation or omission of key facts and 21.3% by selling "unsuitable" investments. The median settlement paid in cases of misconduct amounts to $\$40'000$.

* I am grateful to Sascha Behnk, Uri Gneezy, Manuel Grieder, Ivan Petzev, Joel Sobel, Alexander Wagner, Richard Zeckhauser, and participants at the ESA 2016 World Meeting, the Incentives and Behavior Change conference, and the IMEBESS 2016 for very helpful comments and suggestions. Part of this paper was developed while I was a visiting PhD student at the University of California, San Diego and Harvard University. I thank the Swiss National Science Foundation for financial support.

Similarly, CEOs who can increase the value of their stock holdings by announcing better economic outlooks for their companies can make themselves believe that not all analysts and investors will trust their overly optimistic forecast in order to feel better about the announcement.[1] There are several other fields where asymmetric information leads to unethical behavior and people could engage in self-deception: physicians recommending a suboptimal treatment, car mechanics suggesting unnecessary repairs, and many more.[2]

Mele (1997) argues that these so-called motivated or self-serving beliefs are the key to understanding the dynamics of self-deception. Following Akerlof (1989), who claims that people process information such that they feel comfortable about themselves, economic literature started to analyze how people strategically adjust their first-order beliefs in order to pursue private goals. Babcock and Loewenstein (1997) find self-serving biases in beliefs to play a prominent role in bargaining impasses. People adjust their own perception about what they believe to be a fair outcome depending on the respective private benefits.

This paper studies self-deception, i.e., the usage of motivated beliefs, as a self-serving device to promote unethical advice. Based on experimental data from Amazon Mechanical Turk (mTurk), I analyze whether participants engage in self-deception by distorting their first-order beliefs about the likelihood that their behavior imposes harmful consequences on others. Motivated beliefs might lead to an under- or overestimation of the expected payoff as well as the moral costs.

In a first experiment I consider the case of investment consulting. The investment advisor is aware of the characteristics of several investments available. In addition to a fixed wage for consulting, she might increase her compensation by recommending investment opportunities that are linked to some private side payments. The advisor's recommendation must state which investment opportunity is expected to perform best. As a result, recommending a dominated investment is equal to telling a lie about expected future payoffs.[3] The advisor can hence trade off a higher compensation from recommending suboptimal investments linked to private benefits against the personal harm of misleading the client. In order to provide incentives to lie, I specify the payoffs such that the advisor has incentives not to recommend the best investment opportunity.

In decision stage AUT the advisor faces several investment consulting decisions. As there is no information about the client's behavior the advisor is expected to form her

[1] In fact, Hutton et al. (2012) find that management forecasts are more accurate than analyst forecasts only about 50% of the time.

[2] Kerschbamer and Sutter (2017) provide an extensive meta-study covering lab and field experiments about fraudulent behavior (such as overprovision or overcharging) in markets for credence goods.

[3] For simplicity, there are only two qualitatively different investment opportunities whereas the worse yields a lower expected return for the same risk. As a result, partial lying is not possible and it is simple to determine the objectively better investment.

first-order belief *autonomously*, thereby possibly engaging in self-deception.[4] I elicit her belief in two distinct ways: First, the *stated belief* is determined by directly asking the advisor about her expectation of the client's behavior. Second, I apply Andreoni and Sanchez' (2014) methodology to infer beliefs based on actions. The advisor is asked to choose between the compensation of the investment game where the behavior of the client is not yet known and several lotteries with the same possible payoffs but known probabilities. Based on the characteristics of the lottery for which the advisor starts to prefer the outcome of the game, I infer her *revealed belief*.[5]

In decision stage *EX* the advisor faces the same consulting situations as in AUT but receives *exogenous* information about how often the client will choose the recommended investment. By releasing information about the clients, the advisor is supposed to account for it in her decision making process and to adjust her first-order belief such that it equals the likelihood that a given client chooses the recommended investment. As a consequence, there is no room to self-deceive since exogenous information about the client's behavior is provided. The design of EX allows to infer any advisor's behavior for an extensive range of possible first-order beliefs, i.e., I learn which investment the advisor is going to recommend for any possible information she might have about her client's behavior. Three exogenously induced beliefs are particularly relevant: First, the *actual* rate at which clients choose the investment recommended by the advisor (henceforth *EX:A*), second, the advisor's *stated* belief (*EX:S*), and third, her *revealed* belief (*EX:R*).[6] Comparing the advisor's behavior in AUT, EX:A, EX:S, and EX:R allows to investigate a possible bias in beliefs and its impact on behavior.

The experiment is conducted with a heterogeneous sample of 300 mTurk users in the US. Consistent with prior studies I find that some participants tell the truth even though doing so does not maximize their monetary payoff. In fact, 41.1% of all recommendations made by advisors (in AUT) contain the outperforming investment opportunity. By contrast, 58.9% of all messages recommend a strictly dominated investment opportunity that is linked to private benefits for the advisor. The number of unethical recommendations is significantly increasing in the incentives provided to unethical behavior, i.e., the private payment. These monetary incentives to lie are also called an advisor's economic cost of

[4] The advisor's first-order belief corresponds to her assessment of the likelihood that the client is going to pick the recommended investment opportunity. The higher the subjective probability, the larger is the expected payoff of unethical behavior. However, an increased first-order belief also raises the costs of behaving unethically as misleading the client may induce stronger negative emotions.

[5] The first lottery always pays the highest payoff with certainty. Starting by choosing the lottery at least once is thus a weakly dominant strategy.

[6] In EX:A, the induced belief, i.e., the actual (average) following rate of the clients, is the same for all advisors since it equals the correct belief of an unbiased advisor. Induced beliefs in EX:S and EX:R are individual and correspond to every advisor's elicited stated and revealed belief after AUT.

stating the truth (henceforth ECOST).

Surprisingly, advisors do really poorly in estimating how often clients are going to choose the recommended investment. On average clients follow, i.e., choose the investment opportunity that is recommended by the advisor, in slightly more than 90% of all cases. This is substantially more than advisors' average stated belief (53%) and revealed belief (47%), implying that advisors have biased beliefs.[7] This bias in beliefs leads to distorted behavior such that advisors behave unethically more (or possibly also less) often than if they had perfect knowledge about the clients' behavior. I call this discrepancy *total bias*. The experimental methodology allows to decompose it into two separate components: (i) a *fundamental bias* and (ii) a *self-deceptive bias*.[8]

The fundamental bias is the effect on truthfulness caused by the advisor's general inability to assess the client's actual following rate. It is defined as the difference in advisor behavior if she is informed that clients follow exactly as often as her prior first-order belief (EX:S or EX:R)[9] and her behavior if she knew the correct following rate (EX:A). As a result, the fundamental bias corresponds to the general misperception of client behavior and is computed as the difference EX:R−EX:A, i.e., behavior in EX:R minus behavior in EX:A.

The self-deceptive bias is caused by an advisor who makes use of a motivated belief by adjusting her first-order belief about the client's behavior. In contrast to the fundamental bias, the self-deceptive bias might depend on the private incentives. The self-deceptive bias corresponds to the difference between behavior in AUT where no information about the client's behavior is provided and behavior in EX:R where the induced first-order belief is identical to the elicited revealed belief. Since the advisor's belief in AUT and EX:R is, by construction, similar one would not necessarily expect different behavior. The self-deceptive bias is hence computed as AUT−EX:R. Consequently, the total bias, which is the sum of fundamental and self-deceptive bias, equals AUT−EX:A, i.e., the difference in behavior between AUT where the advisor has no information and EX:A where she possesses correct information about the client's behavior.

I find the absolute effect of the advisors' total bias on lying to be decreasing in ECOST. The higher ECOST, the less biased are advisors and the less does behavior differ. The

[7] Theoretically, advisors could be actively misstating their first-order beliefs, thereby causing a difference between the actual following rate and the stated belief. However, by eliciting revealed beliefs, I control for such misstatements and provide evidence that advisors indeed have biased beliefs.

[8] The experimental methodology herein can also be applied to other settings where people make use of motivated beliefs or, more generally, to any setting that analyzes how a change in the information set alters behavior.

[9] The difference between the stated and the revealed belief corresponds to the misstatement of a given advisor. Since some people have a preference to look more ethical (e.g., Dufwenberg and Gneezy (2000), Andreoni and Bernheim (2009)) people may pretend to do so by manipulating their stated beliefs. I control for possible misstatements by focusing on revealed beliefs instead.

fundamental bias is constant (and nonzero) indicating that advisors anticipate that clients' behavior does not depend on their private benefits. This is reasonable as clients are not informed about the advisor's compensation and the private benefits. The decrease in the total bias is caused by advisors' self-deceptive bias, which is significantly increasing in the magnitude of the private benefits associated with recommending suboptimal investment opportunities.

The data imply that motivated beliefs form a self-deceptive bias that mitigates dishonest investment recommendations if incentives to lie are low and promotes unethical behavior when incentives are high. Exogenously inducing the revealed first-order beliefs, i.e., limiting self-deception, increases lying by 11 percentage points for very small private payments and reduces the likelihood of telling a lie by up to 5 percentage points for high ECOST levels compared to the baseline where advisors form beliefs autonomously.

I also find the biases to be payoff-relevant. For the lowest ECOST level of \$0.20 each advisor loses \$0.03 per decision while the client gains \$0.18 due to self-deception. Increasing ECOST inverts this relationship as the advisor's self-deceptive bias is increasing as well. As a result, at the maximum incentive level of \$1.00, advisors gain \$0.04 and clients lose \$0.03 per decision, respectively. These gains and losses relate to an average payoff per decision of \$1.31 (\$0.85) for advisors (clients) and are hence quite substantial. Given the payoff matrices in this study, the average total surplus (advisor plus client) is U-shaped in ECOST and positive for most incentive levels. Advisors' self-deceptive biases increase combined surplus of advisors and clients. With regards to the fundamental bias, I find that advisors with a fundamental bias lose money no matter in which direction they are biased.

In order to analyze the general validity of the findings in the investment consulting case, I conduct an additional experiment with a neutral framing. I use a sender-receiver game similar to the one studied in Erat and Gneezy (2012) where nature rolls a 6-sided die and the Sender privately observes the outcome from the roll of the die. Knowing the payoffs the Sender must send a message about the actual outcome to the uninformed Receiver. The payments to both players solely depend on whether the Receiver picks the actual outcome or not. Payoffs are similar as in the investment advisory game implying that the Sender is better off if the Receiver does not choose the actual outcome of the die roll, thus providing incentives to lie.

The neutral framing yields slightly more dishonesty but all results regarding biases and engagement in self-deception are confirmed and even more pronounced than in the investment setting. The total bias is also non-linearly decreasing in ECOST because of an increasing self-deceptive bias. Results from the neutral setting confirm that a Sender's payoff is significantly increasing in her self-deceptive bias. Self-deception increases the

Sender's payoff by \$0.12 per decision for ECOST of \$1.00 while the Receiver loses \$0.07. By contrast, the fundamental bias decreases success. To sum up, the findings suggest that people's engagement in self-deception has broad relevance for deception in various environments.

To sum up, this paper makes four major contributions to the literature of self-deception and unethical behavior: First, I provide an experimental methodology to decompose the better-informed party's bias about the behavior of the uninformed into two components: (i) a fundamental bias, which is caused by the inability to assess the other person's behavior, and (ii) a self-deceptive bias caused by motivated beliefs. Second, I find people to engage in self-deception and show the self-deceptive bias to be increasing in the magnitude of private monetary benefits associated with unethical advice. Self-deception proves to be payoff-increasing if the incentives to behave unethically are high and payoff-decreasing if the incentives are low. Third, the fundamental bias does not depend on provided incentives and leads to monetary losses regardless of the direction of the bias. Finally, I can confirm the findings in a specifically framed setting where an investment consultant decides about recommending an underperforming investment opportunity to a client in return for a private payment as well as in a neutral strategic information transmission environment in the sense of Erat and Gneezy (2012). People's engagement in self-deception thus has broad relevance for the study of deception and lying. As a consequence, this paper is able to propose situations in which altering people's information set, e.g, releasing descriptive information about other people's actions, implying that the better-informed party knows how, on average, the less-informed party behaves, might foster ethical behavior.

This study builds on the broad literature that analyzes people's instruments to sustain a proper self-image when behaving unethically (e.g., Akerlof and Kranton (2000), Schweitzer and Hsee (2002), Mazar et al. (2008), Bénabou and Tirole (2011), Chance et al. (2011), Shalvi et al. (2015)) as well as experimental studies on lying and deception in general (e.g., Charness and Dufwenberg (2006), Vanberg (2008), Bucciol et al. (2013), Fischbacher and Heusi (2013)), and in sender-receiver games (e.g., Dickhaut et al. (1995), Blume et al. (1998), Gneezy (2005), Cai and Wang (2006), Sánchez-Pagés and Vorsatz (2007), Wang et al. (2010) among others).

Recently, Gneezy et al. (2015), Di Tella et al. (2015) and Schwardmann and van der Weele (2016) provide experimental evidence that motivated self-deception leads to an increase in unethical behavior. Schwardmann and van der Weele (2016) understand self-deception as overconfidence about one's own performance and show that participants who are incentivized to convince others about their relative performance in fact self-deceive themselves about their own performance and thereby become more successful in persuading others. This paper analyzes self-deception with respect to motivated beliefs about other people's behavior and its impact on unethical advice rather than overconfidence about one's own performance and its effect on the chances of successfully persuading others.

Di Tella et al. (2015) study a corruption game and find people to make use of self-serving biases about the kindness of others in order to behave selfishly. This paper is different in several aspects. First, they analyze people's selfish behavior in a modified dictator game, whereas my focus is on people lying about the true state of the world in a strategic information transmission setting. Second, in their context self-deception considers a self-serving bias regarding the expected kindness of others implying that people convince themselves to be matched with an unkind individual in order to act selfishly without having the feeling of being unfair. I suppose people to engage in self-deception by incorrectly assessing the likelihood that their unethical behavior imposes possibly harmful consequences on others in order to legitimize dishonest reporting. Most importantly my experimental methodology allows to decompose people's biased beliefs into a fundamental and a self-deceptive component while controlling for the fact that some people might be actively misstating their first-order beliefs in the experiment. Thanks to the bias decomposition, I am able to identify people who have fundamentally biased beliefs but are not making use of motivated beliefs even though the effect on behavior could be similar.

Most closely related to this study is Gneezy et al. (2015) who show that advisors distort their judgment of an investment opportunity towards their private incentives if being informed about these incentives in advance. Despite having a similar framing my goal is different. Instead of relating self-deception to the attractivity of the investment opportunities, I focus on the channel, i.e., advisors having biased beliefs about how their clients behave, through which self-deception influences unethical behavior. I find that advisors self-deceive themselves about the likelihood that their clients will choose the recommended investment. Specifically, I show that the increase in unethical behavior is caused by advisors having biased first-order beliefs about their clients' behavior. Since my experimental design allows to directly infer advisor behavior for all possible first-order beliefs, I can then decompose this bias into a fundamental and a self-deceptive part in order to disentangle the net effect of self-deception. In addition, I confirm my findings in a neutral environment and thereby provide evidence that they may also be valid in more general situations where lying takes place and not only in the specific case of investment advisory.

The remainder of the paper is organized as follows. Section 1.2 explains the design of the investment consulting experiment. Section 1.3 describes the sample and its descriptive statistics. Section 1.4 provides results and findings for the investment consulting setting and Section 1.5 for the experiment with a neutral framing. Section 1.6 is a conclusion.

1.2 Experimental Design

1.2.1 Investment Consulting

The first experiment consists of a two-player sender-receiver game in an investment consulting setting. Participants are randomly assigned to either the role of an investment advisor or a client. The participant taking the role of an investment advisor knows the characteristics of all investment opportunities available to her client. There are six investment opportunities available.[10] Without loss of generality, I make three simplifications: (i) there are only two different expected returns, (ii) expected returns are always realized, i.e., the client is paid exactly the expected return, and (iii) all investment opportunities feature the same risk. As there is always one single investment opportunity with a strictly higher expected return, these simplifications make it very easy for the advisor to assess the most favorable investment.[11]

Given the characteristics of the investment opportunities and the compensation schedule the advisor must send a message to the client. The message space consists of six messages:[12]

1. "Investment 1 will yield the highest expected outcome."
2. "Investment 2 will yield the highest expected outcome."
3. "Investment 3 will yield the highest expected outcome."
4. "Investment 4 will yield the highest expected outcome."
5. "Investment 5 will yield the highest expected outcome."
6. "Investment 6 will yield the highest expected outcome."

The investment advisor receives a base salary for consulting the client. In addition, she might increase her payment by recommending investment opportunities that are linked

[10] Since there are six investment opportunities, deception trough telling the truth as propagated by Sutter (2009) should not play an important role. As long as the advisor's first order belief is at least $1/6$, lying is the payoff-maximizing strategy. If it equals $1/6$ the subjective probability to receive the high payoff is $1/6 + 5/6 \times 4/5 = 5/6$ if telling a lie and $1 - 1/6 = 5/6$ if reporting truthfully, respectively. As a result, the two strategies yield the same expected payoff.

[11] In addition, the setting allows for a direct comparison to more general strategic information transmission games such as the one studied in Erat and Gneezy (2012).

[12] Given the wording of the messages behaving unethically is similar to telling a lie.

to private benefits.[13] The client is aware that one investment opportunity outperforms the others from a risk-return perspective but has no additional information about the set of investments. Based on the advisor's recommendation the client consequently picks one investment that determines his payoff as well as the advisor's compensation.

1.2.2 Treatments

The experiment features a within-individual setting. All participants complete all stages. There are two main decision stages and several smaller tasks in the following order:

1. Decision stage I: *Autonomous* belief forming (**AUT**)
2. Belief elicitation I: Stated belief
3. Belief elicitation II: Revealed belief
4. Logical test
5. Decision stage II: *Exogenous* belief induction (**EX**)
6. Questionnaire

In **AUT** the investment advisory game is played for 10 (randomly ordered) rounds with altering payoff matrices.[14] An important parameter is the advisor's first-order belief, i.e., her assessment of the likelihood that the client is going to pick the recommended investment opportunity. In AUT the advisor is expected to form the first-order belief autonomously, as there is no information provided about how clients behave. As a consequence, the advisor might self-deceive herself by distorting her first-order belief.[15] In Appendix 1.A, I provide a simple model that accounts for the possibility to engage in self-deception about the likelihood that the client will choose the recommended investment opportunity while allowing for a broad range of individual-specific moral cost specifications that have been brought up in the literature.

After the 10 decisions in AUT, I elicit the advisor's first-order belief in two different ways. First, the *stated belief* is determined by directly asking the advisor about her expectation of the client's behavior.[16] The stated belief elicitation is unincentivized in order to

[13] In reality, these benefits range from direct commissions for each own product sold, a bonus for reaching a certain amount of product sales from a specific brokerage firm, to a variable compensation part linked to the employer's profit.

[14] Section 1.2.3 discusses the choice of the payoff matrices. They are displayed in Table 1.1.

[15] I assume advisors to maximize the expected utility of lying conditional on being self-deceptively biased rather than to set their level of self-deception strategically in order to maximize utility. As a consequence, self-deception is rather an individual characteristic than a freely selectable parameter.

[16] The exact wording of the stated belief elicitation is: "Out of 100 possible counterparts, how many do you think will follow your messages (i.e., choose the investment you mention in your message)?".

prevent participants from hedging their compensation in the decision stage as suggested by, e.g., Blanco et al. (2010).[17] Second, I apply Andreoni and Sanchez' (2014) methodology to infer beliefs based on actions, thereby controlling for possible misstatements by advisors.

The advisor is asked, how she would prefer to receive her compensation. She is provided a list of 21 lotteries whereas all lotteries are characterized by the same two payoffs as her possible payoffs in the consulting game. For each lottery she has to choose whether she prefers the compensation of the investment game where the behavior of the client is not yet known or the lottery with the same possible payoffs but known probabilities.[18] In fact, each of the 21 lotteries relates one-to-one to an implicit first-order belief between zero and one in the investment advisory game. The first of the 21 lotteries always pays the higher payoff with certainty. Consequently, it is a weakly dominant strategy to start by choosing the lottery.[19] Based on the characteristics of the lottery for which the advisor starts to prefer the outcome of the game, I infer her *revealed belief*. The revealed belief elicitation is framed as a question about how the advisor prefers to receive her compensation. As a consequence, she is not aware that this stage actually elicits beliefs and hence will likely not act in order to be consistent with her behavior in the stated belief elicitation task. Moreover, due to the direct relation to compensation, the advisor is highly incentivized to seriously indicate the marginal lottery for which she starts to prefer the outcome of the game.

The logical test consists of 10 numerical patterns. Participants are asked to complete as many as possible. Each correctly solved pattern adds \$0.10 to the final payment. The time limit for the task is 120 seconds. This task intends to assess the participants' analytical skills, which serve as a proxy for their strategic thinking capability. In addition, the task separates the two treatments in time and prevents advisors from perfectly remembering their behavior in AUT.

In **EX**, the advisor faces the same 10 investment rounds as in AUT but additionally receives a list of 21 possible clients with whom she might be matched. She is told that Client 1 never chooses the investment opportunity that she recommends, Client 2 chooses the recommended investment opportunity in 5 out of 100 cases, Client 3 in 10 out of 100

---

[17] In addition, as shown by Andreoni and Sanchez (2014) people might also misstate first-order beliefs in order to foster their social image by trading off the monetary loss of a worse assessment against decreased moral costs. In order to control for such misstatements, I base my analysis in Section 1.4 and Section 1.5 on revealed beliefs. Results for stated beliefs are provided in Appendix 1.B.

[18] The advisor is also informed about the randomly selected round, the payoff matrix, and her recommendation but not the client's choice.

[19] Only lying advisors with a belief equal to one or honest advisors with a belief equal to zero will be indifferent between the outcome of the game and the lottery that pays the highest payoff with certainty.

cases, ..., and Client 21 in 100 out of 100 cases.[20] By releasing such information about the clients, the advisor is supposed to take it into account by adjusting her belief such that it equals the likelihood that a given client chooses the recommended investment. The advisor is asked to send a recommendation to every possible client without knowing with whom she will be matched. As a consequence, I am able to infer the advisor's behavior for an exhaustive range of exogenously induced first-order beliefs.[21]

Three (out of the 21) induced first-order beliefs receive special attention: (i) the *actual* following rate, i.e., the first-order belief that equals the average likelihood that clients stick to the advisor's recommendation and choose the respective investment (**EX:A**), (ii) the advisor's *revealed* belief (**EX:R**), and (iii) the advisor's *stated* belief (**EX:S**). Note that, while the induced belief in EX:A, i.e., the average following rate of all clients, is the same for all advisors, induced beliefs in EX:S and EX:R correspond to every advisor's individually elicited stated and revealed belief after AUT. By comparing the advisor's behavior for these three induced beliefs, I am able to detect the effect of a possible bias in beliefs on behavior, and more importantly, to decompose this bias into a fundamental (caused by the inability to assess the correct following rate) and a self-deceptive component (caused by motivated first-order beliefs).[22] To the best of my knowledge, this paper is the first to analyze unethical behavior for an extensive range of possible first-order beliefs. The experimental methodology and the respective bias decomposition could furthermore be widely applied to other environments where people make use of motivated beliefs in order to alter behavior.

Finally, a questionnaire eliciting demographic and socio-economic characteristics, protected values for honesty as in Tanner et al. (2009),[23] and prosocial concerns completes

[20] For Client 1, for example, the exact wording of the information provided is: "Participant 1 will choose the investment which you send in your message with 0% probability. Which message would you send Participant 1?". Information about other clients is provided analogously.

[21] Even though I cannot ensure that people actually believe the belief manipulation, there is convincing evidence that they do. First, if they did not believe it they would not take the different induced beliefs into account and behavior in EX would be roughly the same regardless of the induced belief. This is not what the data unravels. Participants' tendency to lie is monotonically increasing in the induced belief. Second, regardless of the incentives provided to report dishonestly, participants on average follow the respective payoff-maximizing strategy given a specific induced belief, indicating that they take the beliefs into account. Figure 1.C.1 in Appendix 1.C displays the advisors' behavior with respect to induced first-order beliefs. Finally, for the sake of this paper, I only require participants to self-deceive less in EX than in AUT and not necessarily that there is no self-deception at all in EX. See the discussion of the degree of self-deception in Appendix 1.A for more details.

[22] While this paper solely focuses on the effect of people's biased beliefs on unethical behavior, one could also analyze the bias in beliefs per se. However, in order to quantify such a bias, I needed to make assumptions about the structural form of every participants' moral costs. I leave this topic to further research.

[23] Gibson et al. (2013) find that this survey-based measure of protected values correlates with preferences for truthfulness.

the experiment.[24] In the end, a lottery anonymously matches advisors and clients in both treatments and randomly determines one round in AUT and one in EX to be relevant for the payment. Participants are never matched with the same counterpart twice. They are paid according to the outcomes in AUT, EX and the performance in the logical test.

Participants taking the role of clients basically follow the same experimental procedure. However, there are two important differences: (i) In AUT, I apply a strategy-method such that clients only decide once which investment to choose given all possible recommendations by the advisor. This is sufficient as clients have no information about the different payoff matrices and hence the strategic decision situation does not vary across rounds. (ii) In EX, clients are randomly assigned to the role of one of the 21 possible clients that are shown to the advisor. As a result, each client is forced to follow the advisor's recommendation in a predetermined number of all cases.

The experiment is programmed and conducted with the experimental software oTree (Chen et al. (2015)). Sessions were run from September to December 2016. The complete set of experimental instructions is provided in Appendix 1.D.

1.2.3 Payoffs

I specify the payoffs such that the investment advisor has private incentives not to recommend the outperforming investment opportunity, thus providing incentives to behave dishonestly. All investment opportunities, which are inferior from a risk-return perspective, are associated with a private payment to the advisor. For simplicity, I assume the private payment to the advisor to be similar for all inferior investments, hence implying two different possible levels of compensation. As there are also two possible return outcomes of the investment opportunities the game features 2×2 payoff matrices. As a result, partial lying is not possible and hence there are no different sizes of lies.

I alternate the incentives to lie, i.e., the private payment associated with inferior investment opportunities from round to round producing different strategic decision situations for the investment advisor.[25] There are five differing values of ECOST ranging from $0.20 to $1.00.[26] Each ECOST level is used in two out of the 10 randomly ordered rounds. Table 1.1 displays all payoff matrices.

[24] I use a simple survey-based prosocial concerns measure as in Gibson et al. (2013). There are more sophisticated prosocial concerns measures, e.g., Van Lange et al. (1997), which could be used if one wanted to focus especially on the effects of prosociality.

[25] In addition, the possible outcomes to advisor and client are slightly adjusted in order to keep the difference in payoffs between both players constant in a given state. This allows me to control for any effects of inequity aversion.

[26] The effect on the hourly wage is substantial. For example, ECOST of $1.00 translates to an increase in the hourly wage of $12.00 as both stages took participants approximately 5 minutes to complete.

Table 1.1: Payoff matrices

ECOST	$\left(\Pi^A, \Pi^C\right)$ good investment	$\left(\Pi^A, \Pi^C\right)$ bad investment	$\left(\Pi^A, \Pi^C\right)$ good investment	$\left(\Pi^A, \Pi^C\right)$ bad investment
\$0.20	(\$1.00, \$1.70)	(\$1.20, \$0.30)	(\$1.10, \$1.80)	(\$1.30, \$0.40)
\$0.40	(\$0.90, \$1.60)	(\$1.30, \$0.40)	(\$1.00, \$1.70)	(\$1.40, \$0.50)
\$0.60	(\$0.80, \$1.50)	(\$1.40, \$0.50)	(\$0.90, \$1.60)	(\$1.50, \$0.60)
\$0.80	(\$0.70, \$1.40)	(\$1.50, \$0.60)	(\$0.80, \$1.50)	(\$1.60, \$0.70)
\$1.00	(\$0.60, \$1.30)	(\$1.60, \$0.70)	(\$0.70, \$1.40)	(\$1.70, \$0.80)

This table summarizes all 10 payoff matrices. The tuple $\left(\Pi^A, \Pi^C\right)$ denotes the advisor's and the client's payoff, respectively. In the experiment, payoffs are displayed in experimental currency units. However, they can also be interpreted as cents since one experimental currency unit in the experiment corresponds to \$0.01 real money. ECOST is not shown to participants. The client may either choose the good investment opportunity or the bad investment opportunity. Selecting the bad investment opportunity will trigger private payments to the advisor and hence lead to a higher advisor payoff.

As an illustration, one specific payoff matrix, which corresponds to the first matrix depicted in the top row in Table 1.1, is displayed in Figure 1.1.

	Client: good investment	Client: bad investment
Advisor	\$1.00 / \$1.70	\$1.20 / \$0.30

Figure 1.1: Payoff matrix example

In this particular situation the advisor's base compensation is \$1.00. In addition, she might achieve a private payment of \$0.20 (= \$1.20 − \$1.00) if the client chooses the bad investment opportunity. For the client the good investment opportunity will yield \$1.70 and the bad investment opportunity \$0.30. Hence, the payoffs are (\$1.00, \$1.70) if the client picks the good investment opportunity, i.e., \$1.00 for the advisor and \$1.70 for the client and (\$1.20, \$0.30) if he picks the bad investment opportunity, i.e., \$1.20 for the advisor (\$1.00 plus a private payment of \$0.20) and \$0.30 for the client. Consequently, the monetary incentives to behave unethically and to recommend a bad investment opportunity amount to \$0.20.

1.3 Data

I recruited 300 US participants on the crowdsourcing internet marketplace mTurk. In order to guarantee participant validity, I require subjects to have successfully completed at least 1000 human intelligence tasks (HITs) on mTurk with an overall approval rate of more than 95%.[27] In addition, there are two understanding questions. 79% of all participants solved the first understanding question correctly and 81% were able to answer the second question adequately.[28]

The sample consists of 51% men and 49% women. The sample features an age range from 18 to 68 years and a variety in highest completed education from elementary to graduate school. The median participant is 31 years old, holds an undergraduate degree and has a household income of $50′000.[29]

Participants' earnings were a $0.25 show-up fee plus an average bonus of about $2.75 for advisors and $1.65 for clients. Given that the experiment took about 15 minutes for advisors and 10 minutes for clients, advisors earned $12.00 ($= 3.00/15 \times 60$) per hour on average and clients $11.40 ($= 1.90/10 \times 60$), respectively. Table 1.2 summarizes the data's descriptive statistics. The distribution of protected values and prosocial concerns is consistent with Tanner et al. (2009) and Gibson et al. (2013).

The elicitation of participants' revealed beliefs requires a clean-up of the data. First, participants who switch multiple times between preferring the lottery and the outcome of the game or start with preferring the outcome of the game and then switch to the lottery are excluded as these strategies are strictly dominated and cannot be related to a specific revealed belief.[30] Second, participants who always prefer either of the two options are excluded as this would imply a belief of zero or one. However, it could also indicate a lack of understanding or attention. As a precaution, all participants who always prefer either of the two options, except those that also state a belief of zero or one,[31] are dropped.[32] In total 54 participants are excluded. Including these participants would confirm and even strengthen my results.

---

[27] There are several studies showing that mTurk experiments yield consistent results to common laboratory settings even though the stakes are smaller (e.g., Paolacci et al. (2010), Horton et al. (2011), Amir et al. (2012)).

[28] Despite that the training questions are not solved perfectly, the results are robust to the exclusion of those who did not solve them correctly.

[29] Bohannon (2016) argues that, despite the heterogeneity of the sample, many participants on mTurk are professional experiment takers just as in many laboratory environments where participants are hired out of a pool of regular experiment takers.

[30] Participants who commit one single mistake, i.e., switching to the other option for exactly one lottery and then immediately switching back, are included and the respective action is reversed. My findings are not sensitive to an exclusion of these participants.

[31] In fact, there is only one participant who states a first-order belief of zero and one who reports a belief of one.

[32] Always preferring the lottery could indicate that an advisor is ambiguity averse and prefers to enter a lottery with known probabilities despite a possibly smaller expected payoff. Similarly, participants with a preference for ambiguity might always prefer to be paid the outcome of the game. Excluding these participants prevents the measure of revealed beliefs to be distorted by such ambiguity aversion.

Table 1.2: Descriptive statistics - Investment consulting

	Mean	Std.	Min	Max	Obs.
Sex	0.51	0.50	0	1	300
Age	33.4	9.83	18	68	300
Education	2.67	0.68	1	4	300
Income	2.90	1.46	0	6	300
Religion	0.42	0.49	0	1	300
PV	4.46	1.16	1.67	7.00	300
Prosoc	3.44	1.83	1.00	7.00	300
Stated belief	0.53	0.25	0.00	1.00	150
Revealed belief	0.46	0.25	0.00	0.95	96
Numskill	1.5	1.16	0	6	300
Bonus	2.29	0.78	0.6	3.9	300

This table shows descriptive statistics for all advisors and clients. Sex is a binary variable taking the value one if the participant reports to be male, and zero if female. Education corresponds to the highest completed education stage and takes the following values: 0: No schooling, 1: Elementary school, 2: High school, 3: Undergraduate degree, 4: Graduate school. Income reflects yearly household income and takes the following values: 1: $< \$20'000$, 2: $\$20'000 - 39'999$, 3: $\$40'000 - 59'999$, 4: $\$60'000 - 79'999$, 5: $\$80'000 - 99'999$, 6: $> \$100'000$. Religion denotes a binary variable which takes the value one if the participant indicates to belong either to Christianity, Islam, Judaism, Hinduism, Buddhism or a Folk religion, and zero else. PV reflects an index of protected values for honesty borrowed from Tanner et al. (2009). Prosoc is an index capturing prosocial concerns. The full set of questions is depicted in Appendix 1.D. Stated belief is the advisor's self-reported first-order belief about the client's behavior. Revealed belief reflects the action-implied first-order belief according to the methodology of Andreoni and Sanchez (2014). Numskill is the number of correctly solved numerical patterns in the logical test. Bonus corresponds to the variable payoff achieved.

1.4 Results

1.4.1 First-order Beliefs

I start by analyzing the distributions of the two first-order belief measures and their implicit relationship. Figure 1.2 presents histograms for stated and revealed beliefs, whereas the latter are derived according to Andreoni and Sanchez' (2014) methodology. The distribution of stated beliefs features a peak at 1/2, which might be a focal point for many advisors. The average stated belief equals 55% and is in line with those elicited by Rode (2010) (51%) in a standard communication game.[33] Revealed beliefs average 47% and feature a smoother distribution. Figure 1.3 links the two measures in a scatter plot.

The overall relationship between stated and revealed beliefs is positive but the correlation is small (0.07). However, there is a substantial difference between participants

[33] Considering all participants, i.e., not eliminating participants with invalid revealed beliefs, the average stated belief equals 53% (see Table 1.2).

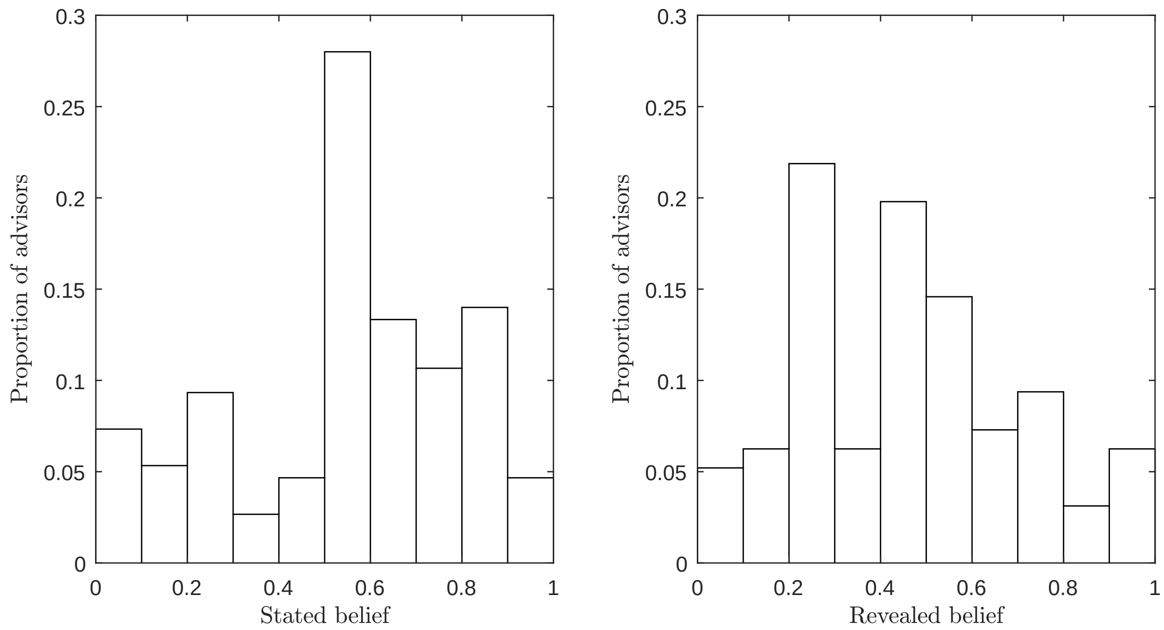


Figure 1.2: Histograms of stated (left) and revealed advisor beliefs (right). Stated beliefs are unincentivized and self-reported. Revealed beliefs are derived according to the methodology of Andreoni and Sanchez (2014).

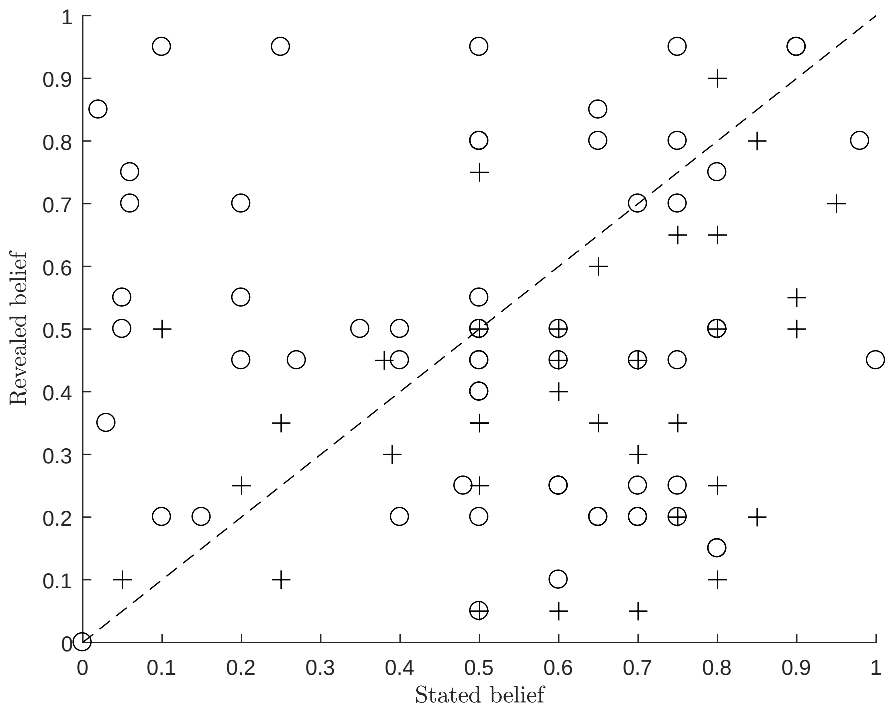


Figure 1.3: Stated versus revealed advisor beliefs. Plus markers indicate participants who lie in at most 50% of all decisions. Circles mark participants who lie in more than 50% of all decisions.

who are rather honest and those who are not. While the correlation between stated and revealed beliefs of participants who lie in more than 50% of all cases is 0.02, it is 0.34 for those who lie in 50% or less cases. Most honest participants overstate their stated beliefs such that they end up below the 45-degree line in Figure 1.3. One motive is that these individuals want to look particularly ethical by telling the truth because they claim that they expect others to believe their recommendation (e.g., Dufwenberg and Gneezy (2000), Andreoni and Bernheim (2009)).

Many dishonest participants, by contrast, underreport their belief and act as if they thought that their unethical behavior would not hurt clients very often. However, some even report values for which lying would no longer be the payoff-maximizing strategy, i.e., if the first-order belief is smaller than 1/6. Some dishonest advisors also overstate their belief because they possibly want to highlight their capability to understand that lying becomes more profitable the higher is their respective belief.

Advisors do really poorly in estimating how often clients are going to choose the investment recommended. On average, clients stick to the advisor's recommendation in slightly more than 90% of all cases. 128 (of 150) clients are always credulous, 4 clients never follow, and 18 elect the recommended investment sometimes. This is substantially more than what advisors believe (regardless whether measured by stated or revealed beliefs), hence indicating that advisors have biased beliefs.

1.4.2 Unethical Behavior

Consistent with studies conducted in the laboratory, I find participants to tell the truth even though it does not maximize their monetary payoff. In fact, 41.1% of all recommendations made by advisors in AUT contain the investment opportunity that really outperforms all others. 58.9% of all messages recommend a dominated investment opportunity that is linked to private benefits for the advisor. The share of these unethical recommendations is increasing in ECOST.[34] Figure 1.4 displays the relationship between ECOST and unethical advice in AUT.

Table 1.3 lists how demographics and socio-economic characteristics influence unethical investment recommendations in AUT. Lying most importantly depends on the incentives to behave unethically, i.e., the private benefits from recommending an underperforming investment opportunity. The proportion of lies is significantly increasing in ECOST and decreasing in ECOST^2. This is consistent with Figure 1.4 which illustrates that the increase in lying is diminishing for higher levels of ECOST. Remarkably, neither stated nor revealed beliefs explain a significant share of unethical behavior, indicating

[34] This is consistent with Gneezy (2005) in another sender-receiver game and also Gibson et al. (2013) in a non-strategic setting.

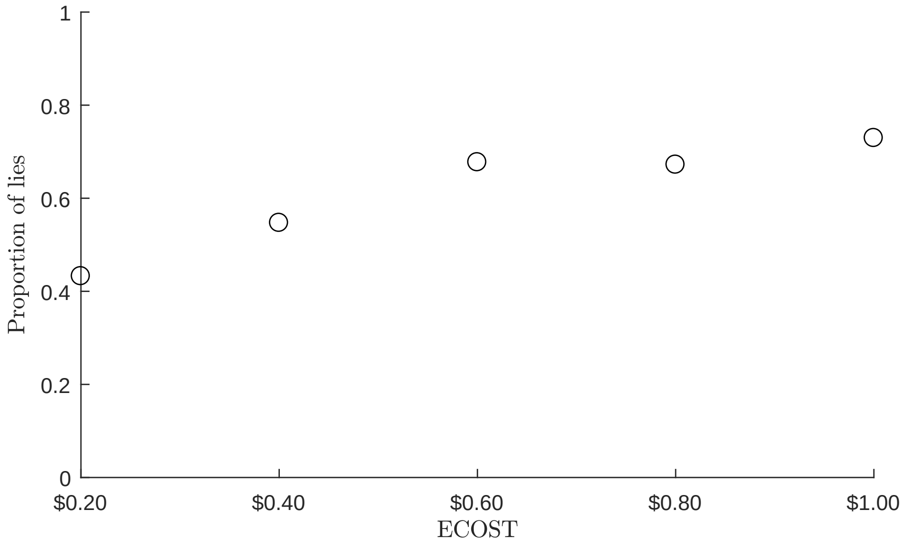


Figure 1.4: Unethical investment recommendations across ECOST levels in AUT. Each circle corresponds to the average proportion of lies in two situations with similar ECOST.

that beliefs affect both the advisor's expected monetary payoff and her moral costs of behaving unethically. If moral costs were independent of first-order beliefs, lying would be significantly increasing in the respective beliefs as the expected monetary payoff of lying grows in the advisor's subjective assessment of the likelihood that the client will choose the recommended investment.

The effect of demographic and socio-economic characteristics on lying is limited. Men lie insignificantly more. Lying is also increasing in the advisor's age. In contrast, religious and well-educated participants tell the truth more often. The household income's effect on lying is ambiguous. Moreover, the higher a participant's logical skills, as measured by the number of numerical patterns solved, the lower is the proportion of lies. Finally, preferences for protected values for honesty are able to explain a share of the lying decisions. The stronger an advisor's protected values, the less likely she is going to recommend a bad investment opportunity.

1.4.3 Bias Decomposition

The investment advisors' assessment of their clients' behavior in AUT is biased. Regardless of the belief measure they expect clients to choose the recommended investment opportunity less than what clients actually do. This bias in beliefs leads to a bias in behavior implying that advisors behave unethically more (or conceivably also less) often than they would if they had perfect knowledge about the clients' behavior. I call this bias *total bias*. It is determined by comparing behavior in AUT to behavior in EX:A when

Table 1.3: Determinants of unethical advice in investment consulting

Dependent variable:			Lie			
	(1)	(2)	(3)	(4)	(5)	(6)
ECOST	1.493***	4.989***	1.499***	4.962***	1.754***	7.400***
	(0.235)	(1.892)	(0.236)	(1.883)	(0.311)	(2.737)
ECOST^2		-1.810***		-1.819***		-1.714**
		(0.632)		(0.634)		(0.810)
Stated belief			-0.364	-0.360		
			(0.376)	(0.378)		
Revealed belief					0.260	0.272
					(0.606)	(0.615)
Sex	0.345	0.349	0.301	0.305	0.389	0.396
	(0.264)	(0.265)	(0.266)	(0.267)	(0.348)	(0.350)
Age	0.016	0.017	0.015	0.015	0.003	0.004
	(0.016)	(0.016)	(0.016)	(0.016)	(0.021)	(0.021)
Education	-0.101	-0.100	-0.105	-0.104	-0.063	-0.062
	(0.216)	(0.217)	(0.217)	(0.218)	(0.251)	(0.252)
Income	-0.122	-0.123	-0.122	-0.123	0.040	0.040
	(0.108)	(0.108)	(0.109)	(0.109)	(0.128)	(0.129)
Religion	-0.305	-0.306	-0.296	-0.297	-0.418	-0.421
	(0.257)	(0.258)	(0.258)	(0.259)	(0.322)	(0.324)
PV	-0.590***	-0.434**	-0.599***	-0.451**	-0.718***	-0.362
	(0.165)	(0.181)	(0.161)	(0.181)	(0.223)	(0.252)
ECOST × PV		-0.270		-0.258		-0.629*
		(0.267)		(0.266)		(0.375)
Prosoc	-0.060	-0.029	-0.065	-0.032	-0.080	0.067
	(0.088)	(0.116)	(0.088)	(0.115)	(0.112)	(0.169)
ECOST × Prosoc		-0.056		-0.060		-0.264
		(0.179)		(0.177)		(0.251)
Numskill	-0.011	-0.011	-0.021	-0.021	-0.034	-0.035
	(0.124)	(0.125)	(0.123)	(0.124)	(0.155)	(0.156)
Constant	2.437**	1.167	2.740***	1.489	2.706**	0.171
	(0.989)	(1.244)	(1.040)	(1.316)	(1.256)	(1.714)
Observations	1,500	1,500	1,500	1,500	960	960
Number of participants	150	150	150	150	96	96
(Pseudo) R^2	0.097	0.100	0.099	0.102	0.119	0.126

This table presents coefficients of logit regressions in AUT. The dependent variable is the binary variable Lie which takes the value one if the advisor behaves unethically, and zero else. ECOST denotes economic costs of stating the truth, i.e., the monetary incentives to recommend a bad investment opportunity. Stated belief reflects the advisor's self-reported first-order belief about the likelihood that the client will choose the recommended investment. Revealed belief denotes the advisor's first-order belief as implied by the methodology of Andreoni and Sanchez (2014). Sex is a binary variable taking the value one if the participant reports to be male, and zero if female. Education takes the following values: 0: No schooling, 1: Elementary school, 2: High school, 3: Undergraduate degree, 4: Graduate school. Income reflects yearly household income and takes the following values: 1: $< \$20'000$, 2: $\$20'000 - 39'999$, 3: $\$40'000 - 59'999$, 4: $\$60'000 - 79'999$, 5: $\$80'000 - 99'999$, 6: $> \$100'000$. Religion denotes a binary variable which takes the value one if the participant indicates to belong either to Christianity, Islam, Judaism, Hinduism, Buddhism or a Folk religion, and zero else. PV reflects an index of protected values for honesty. Prosoc captures prosocial concerns. Numskill is the number of correctly solved numerical patterns in the logical test. Robust standard errors are obtained by clustering at the individual level and appear in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

the correct belief (90.3%), which corresponds to the actual following rate, is induced.[35] Figure 1.5 displays advisors' average total bias across ECOST.

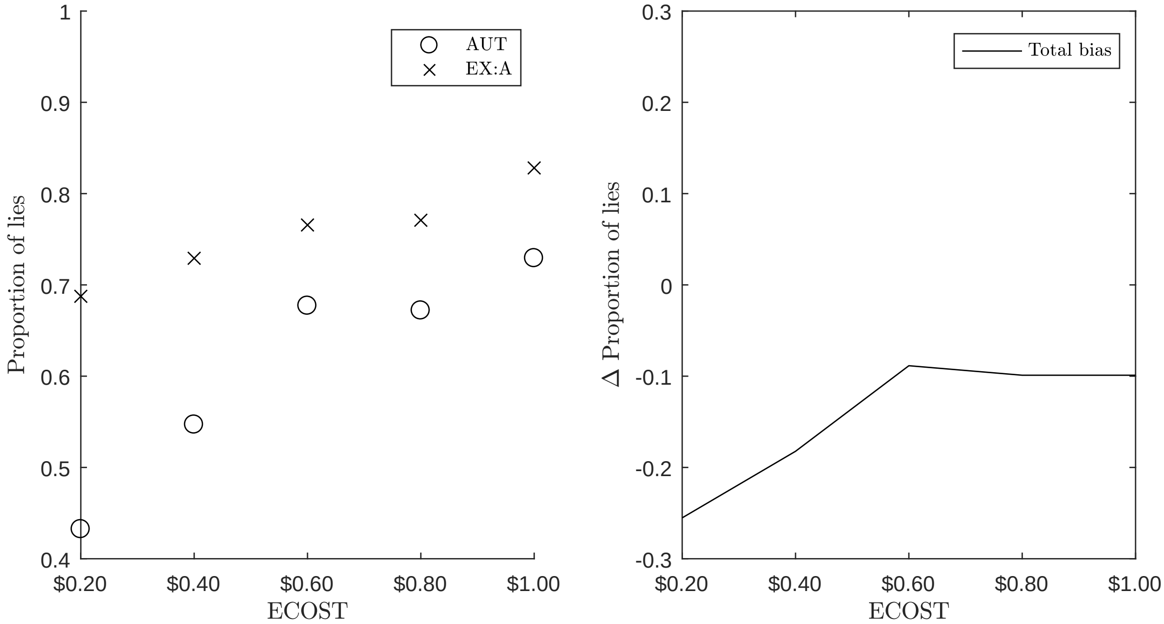


Figure 1.5: Unethical behavior in AUT and EX:A (left), and advisors' total bias in investment consulting (right). Circles indicate the proportion of lies for given ECOST in AUT. Crosses correspond to the proportion of lies in EX:A when the actual following rate (90%) is exogenously induced. The total bias equals the difference between the proportion of lies in AUT and EX:A.

If advisors know the actual following rate, recommending a bad investment opportunity becomes more popular as it is very likely that the client will choose the recommended investment. Hence, Figure 1.5 provides evidence that investment advisors' assessment of the clients' behavior diverges from their real behavior since the former would have no incentive to behave differently in AUT compared to EX:A if they had the same expectation. I observe the absolute level of the total bias to be decreasing in ECOST. The higher ECOST, the less biased are the advisors and the less does behavior between the two treatments differ. Overall the total bias is substantial, decreasing from 25 percentage points if ECOST equals \$0.20 to 10 percentage points if ECOST amounts to \$1.00.

Thanks to the experimental methodology, I am able to decompose the total bias into two separate components: (i) a *fundamental bias* and (ii) a *self-deceptive bias*. The fundamental and the self-deceptive bias always add up to the total bias.

The *fundamental bias* is the effect on truthfulness caused by an advisor's general inability to assess the client's actual following rate. Technically speaking, it corresponds

[35] Since I only have behavioral data for 21 different induced beliefs, I use the induced belief that is closest to the actual following rate. The closest available induced belief is 90%.

to the difference between behavior in EX:R where the induced first-order belief equals her prior first-order belief and EX:A where it corresponds to the actual following rate. Since the client does not know the advisor's compensation and especially the private benefits, the fundamental bias should be independent of the incentives provided to the advisor.

The *self-deceptive bias* is caused by an advisor's use of motivated beliefs. This advisor engages in self-deception by adjusting her first-order belief about the client's behavior. In contrast to the fundamental bias, the self-deceptive bias might depend on the incentives to behave unethically. Stronger incentives might tempt the advisor to bias beliefs in order to restructure her cognition. The self-deceptive bias corresponds to the difference between behavior in AUT when no information about the clients' behavior is provided and behavior in EX:R where the previously elicited first-order belief is induced.[36] If advisors did not engage in self-deception, behavior in AUT would be expected to be similar to behavior in EX:R as advisors have the same beliefs about their clients' behavior.[37]

Figure 1.6 displays the proportion of lies in AUT where advisors might self-deceive, EX:A where the correct following rate is stated, and EX:R where each advisor's revealed belief is induced. The total bias equals AUT−EX:A, the fundamental bias EX:R−EX:A, and the self-deceptive bias reflects AUT−EX:R.

Figure 1.6 shows that advisors' motivated beliefs lead to a negative self-deceptive bias if incentives to behave unethically are low and positive if incentives are high. While a negative self-deceptive bias implies that the advisor mitigates dishonest investment recommendations in AUT, a positive bias leads to more unethical behavior when forming beliefs autonomously. Exogenously inducing the revealed first-order beliefs, i.e., limiting self-deception, increases lying by 11 percentage points for the lowest private payment and reduces the likelihood of telling a lie by up to 5 percentage points for high ECOST levels compared to AUT where advisors can engage in self-deception.

As the fundamental bias (the general misperception of client behavior) is roughly constant, the absolute decrease in the total bias is caused by advisors' self-deceptive biases, which decrease lying in AUT for low levels of ECOST and increase lying for high

---

[36] E.g., in case advisor A revealed in the belief elicitation that she expects her client to follow with 60% probability in AUT, I compare her behavior in AUT with her behavior in EX:R, where the client will actually follow with 60% probability.

[37] To control for some advisors possibly strategically misstating their beliefs, I base the analysis on revealed beliefs. However, all the subsequent findings also hold if one considered behavior in EX:S where stated first-order beliefs were induced instead. Even though the correlation between stated and revealed beliefs is low, both measures provide qualitatively similar findings. The low correlation suggests that participants are not strategically misstating their beliefs in order to obscure their self-deceptive tendencies in which case I could not observe a self-deceptive bias by analyzing stated beliefs. In addition, most participants' stated and revealed beliefs are both far off the actual following rate, indicating that also the fundamental bias persists. All results and findings based on stated beliefs are displayed in Appendix 1.B.

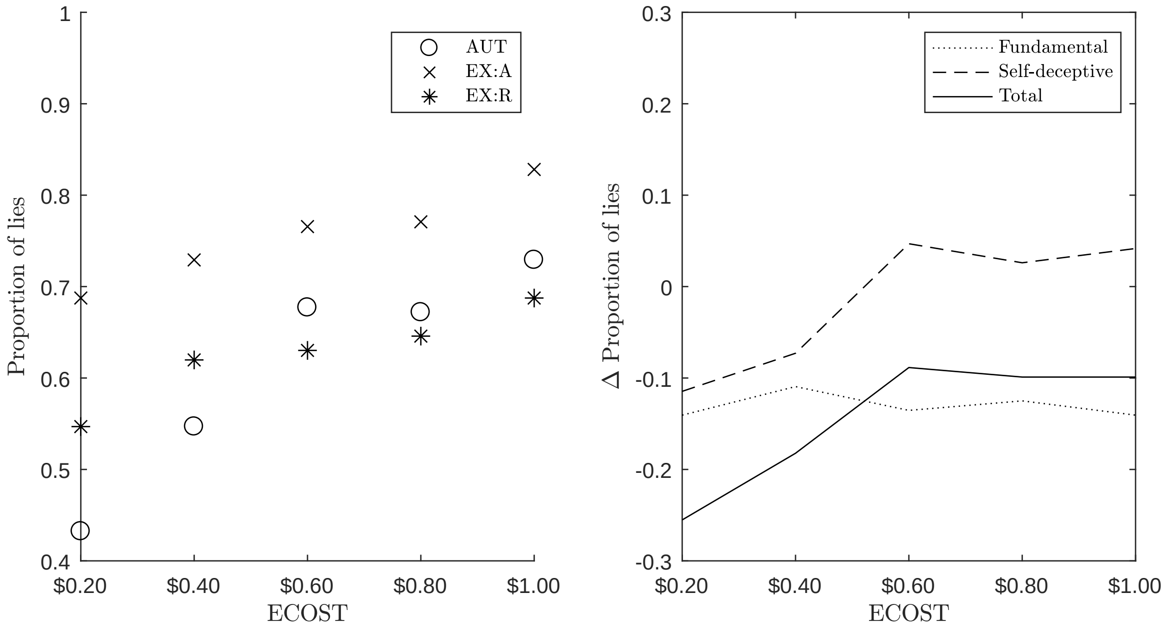


Figure 1.6: Unethical behavior in AUT, EX:A and EX:R (left), and the bias decomposition in investment consulting (right). Circles indicate the proportion of lies for given ECOST in AUT. Crosses correspond to the proportion of lies in EX:A, and asterisks denote the proportion of lies in EX:R. The fundamental bias is computed as EX:R−EX:A. The self-deceptive bias corresponds to AUT−EX:R. The total bias equals the sum of fundamental and self-deceptive bias.

levels of ECOST. Having a constant fundamental bias indicates that advisors anticipate that the clients' behavior does not depend on private benefits. This is reasonable as clients are neither informed about the advisor's compensation nor the private benefits.

Table 1.4, which analyzes the determinants of the total bias, the fundamental bias, and the self-deceptive bias, empirically confirms the indicative evidence provided in Figure 1.6 that the absolute level of the total bias is decreasing in ECOST because advisors' self-deceptive bias is significantly increasing in the value of the private benefits associated with recommending underperforming investment opportunities. A $0.10 increase (at the mean) of the incentives provided to recommend a bad investment project leads to an increase in the self-deceptive bias of 2.1 ($= 0.1 \cdot (0.619 - 2 \cdot 0.344 \cdot 0.6)$) percentage points ceteris paribus. This translates one-to-one to a 2.1 percentage point increase in the proportion of lies even without considering the effect of an incentive increase on lying per se.

In addition, the absolute level of the fundamental bias is shown to be positive and independent of private incentives. Advisors with strong protected values for honesty and prosocial concerns are slightly less fundamentally biased than their peers. Socio-economic and demographic characteristics hardly have an effect on the magnitude of the biases except that religious people are significantly more (negatively) biased in total.

Table 1.4: Determinants of the bias decomposition in investment consulting

Dependent variable:	Total bias		Fundamental bias		Self-deceptive bias	
	(1)	(2)	(3)	(4)	(5)	(6)
ECOST	0.734***	1.208***	0.115	0.105	0.619**	1.103**
	(0.263)	(0.413)	(0.230)	(0.455)	(0.284)	(0.485)
$ECOST^2$	-0.446**	-0.446**	-0.102	-0.102	-0.344	-0.344
	(0.197)	(0.197)	(0.167)	(0.167)	(0.214)	(0.215)
Sex	0.007	0.007	0.037	0.037	-0.030	-0.030
	(0.053)	(0.053)	(0.082)	(0.082)	(0.088)	(0.088)
Age	0.002	0.002	0.004	0.004	-0.002	-0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Education	-0.048	-0.048	-0.046	-0.046	-0.002	-0.002
	(0.037)	(0.037)	(0.052)	(0.052)	(0.056)	(0.056)
Income	0.018	0.018	0.021	0.021	-0.003	-0.003
	(0.019)	(0.019)	(0.027)	(0.027)	(0.032)	(0.032)
Religion	-0.123**	-0.123**	0.019	0.019	-0.142	-0.142
	(0.056)	(0.056)	(0.089)	(0.089)	(0.092)	(0.092)
PV	0.039	0.080*	0.065*	0.068	-0.026	0.011
	(0.026)	(0.043)	(0.036)	(0.042)	(0.041)	(0.047)
ECOST × PV		-0.067		-0.005		-0.062
		(0.048)		(0.041)		(0.052)
Prosoc	0.006	0.040	0.044	0.038	-0.038	0.002
	(0.014)	(0.028)	(0.028)	(0.036)	(0.029)	(0.037)
ECOST × Prosoc		-0.057*		0.010		-0.067
		(0.033)		(0.036)		(0.044)
Numskill	-0.026	-0.026	-0.008	-0.008	-0.017	-0.017
	(0.023)	(0.024)	(0.031)	(0.031)	(0.036)	(0.036)
Constant	-0.484**	-0.769**	-0.656**	-0.650*	0.172	-0.119
	(0.202)	(0.299)	(0.254)	(0.328)	(0.262)	(0.315)
Observations	960	960	960	960	960	960
Number of participants	96	96	96	96	96	96
R-squared	0.051	0.054	0.044	0.044	0.034	0.036

This table presents coefficients of OLS regressions. The dependent variables are the total bias in columns (1) and (2), the fundamental bias in columns (3) and (4), and the self-deceptive bias in columns (5) and (6). The total bias is computed as AUT−EX:A, the fundamental bias as EX:R−EX:A, and the self-deceptive bias as AUT−EX:R. ECOST denotes economic costs of stating the truth, i.e., the monetary incentives to recommend a bad investment opportunity. Sex is a binary variable taking the value one if the participant reports to be male, and zero if female. Education takes the following values: 0: No schooling, 1: Elementary school, 2: High school, 3: Undergraduate degree, 4: Graduate school. Income reflects yearly household income and takes the following values: 1: $< \$20'000$, 2: $\$20'000 - 39'999$, 3: $\$40'000 - 59'999$, 4: $\$60'000 - 79'999$, 5: $\$80'000 - 99'999$, 6: $> \$100'000$. Religion denotes a binary variable which takes the value one if the participant indicates to belong either to Christianity, Islam, Judaism, Hinduism, Buddhism or a Folk religion, and zero else. PV reflects an index of protected values for honesty. Prosoc captures prosocial concerns. Numskill is the number of correctly solved numerical patterns in the logical test. Robust standard errors are obtained by clustering at the individual level and appear in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

Figure 1.7 provides the cross-section of self-deceptive biases within the population aggregated over all decisions (in AUT) and incentive levels. I find the distribution to peak at zero and to have a slightly negative mean. On average 30% of the advisors do not feature self-deceptive tendencies.

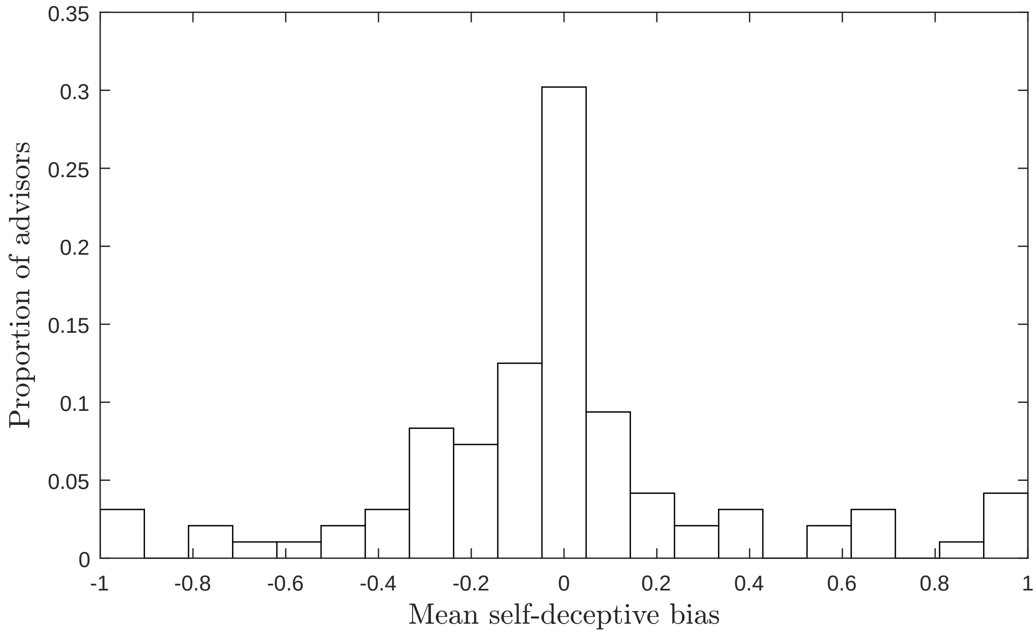


Figure 1.7: Cross-section of self-deceptive biases in investment consulting. The self-deceptive bias is computed as the difference between behavior in AUT and EX:R averaged over all ECOST levels.

In order to further investigate participants' engagement in self-deception, I now analyze the effect of restricting self-deception. I compare lying in AUT where self-deception is possible to lying in EX:R where it is limited. Table 1.5 provides evidence that the induction of every advisor's individual revealed belief increases lying for low ECOST and decreases lying for high levels of ECOST. In other words, advisors engage in self-deception in AUT if they are allowed to form their first-order beliefs about clients autonomously. As a consequence of self-deception, they recommend bad investments that are linked to private benefits less often if the private incentives are low but more if the incentives to do so are high compared to the situation where they are informed that clients will choose the recommended investment opportunity exactly as often as they think.

From a social perspective it might be interesting to analyze whether providing better-informed parties with information about descriptive average behavior of the less-informed party is welfare enhancing. Providing this would be easy and cost-efficient in many settings. However, whether exogenously inducing the correct first-order belief per se might reduce lying depends on the sign of the total bias. If the total bias is positive, making the actual following rate of the uninformed party public will reduce unethical behavior overall. This might be the case either if the fundamental bias is positive or if the self-deceptive bias is positive. In this investment consulting setting the effect of limiting self-deception (self-deceptive bias) would be dominated by the effect of people correcting their beliefs for the fundamental bias once accurate descriptive information about clients'

Table 1.5: Limiting self-deception in investment consulting

Dependent variable:		Lie	
	(1)	(2)	(3)
ECOST	2.600***	3.079***	4.779**
	(0.828)	(0.867)	(2.035)
ECOST^2	-1.164**	-1.151*	-1.047*
	(0.593)	(0.594)	(0.570)
EX:R	0.070	0.639***	0.632***
	(0.196)	(0.226)	(0.219)
ECOST × EX:R		-0.978***	-0.977***
		(0.301)	(0.307)
Controls	Yes	Yes	Yes
Interactions	No	No	Yes
Observations	1,920	1,920	1,920
Number of participants	96	96	96
(Pseudo) R^2	0.090	0.093	0.095

This table presents coefficients of logit regressions. The dependent variable is the binary variable Lie which takes the value one if the advisor behaves unethically, and zero else. ECOST denotes economic costs of stating the truth, i.e., the monetary incentives to recommend a bad investment opportunity. EX:R is a binary variable variable which takes the value one in EX:R where every advisor's revealed belief is induced. Controls include Sex, Age, Education, Income, Religion, PV, Prosoc, and Numskill. Interactions indicates interaction terms of PV and Prosoc with ECOST. Robust standard errors are obtained by clustering at the individual level and appear in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

behavior was released. Despite the positive self-deceptive bias for high ECOST levels, the fundamental bias is still too pronounced implying the total bias to be negative. Hence, making descriptive information about average behavior of clients public still leads to an increase in deceitful investment recommendations. However, this must not necessarily be the case in other environments where people behave unethically. This study suggests that releasing descriptive information about the actual behavior of the uninformed party to have a positive effect on truth-telling whenever the better-informed party's fundamental bias is not too big and incentives to behave unethically are substantial.

1.4.4 Payoff Implications

Does advisors' engagement in self-deception increase their payoff? There are two main drivers that might trigger advisors to use motivated beliefs: on the one hand, the chance of a higher expected payoff and, on the other hand, smaller ethical concerns. While I am able to measure the likelihood of a higher payoff, I cannot directly assess moral costs without postulating assumptions about the functional specification. As a result, I focus on analyzing the effect of self-deception on compensation. I measure the effect on economic success in two ways: (i) a binary variable (High Payoff AUT) that captures whether the advisor ended up in the high payoff state, and (ii) the advisor's behavior-implied compensation in AUT (Payoff AUT). Whereas (i) controls for the specifics of the herein considered payoff matrices, (ii) allows to infer the absolute effect on monetary compensation.

Table 1.6 provides evidence that engaging in self-deception significantly increases advisors' chance of ending up in the most favorable state and hence also their payoff. The effect of the self-deceptive bias on people's payoff is positive and slightly concave. For advisors with a positive self-deceptive bias, an additional increase of their bias implies a smaller payoff gain than the same increase for an unbiased advisor. For advisors experiencing a negative self-deceptive bias, however, further enlarging their bias is costly.

With regards to the fundamental bias, I find the coefficient on the squared term to be negative and statistically significant implying that advisors with a non-zero fundamental bias lose money regardless of the direction of the bias. The greater the absolute magnitude of the fundamental bias, the lower is the achieved payoff. Similarly, advisors with strong protected values for honesty are earning significantly less because they report honestly more often. An increase of one unit in the 7-point Likert scale leads to a payoff reduction of up to $0.10.

Next, I analyze the effective impact of self-deception on participants' payoffs. I match advisor behavior in EX:R, where self-deception is limited, to client behavior in AUT and derive the resulting payoffs. As a consequence, I can compare the payoffs of advisors and clients in AUT to the payoffs they would have achieved if the advisors played the same decisions without any self-deceptive bias. The resulting differences in payoffs and the different chances of ending up with a higher payoff per se are shown in Figure 1.8.

The left-hand side of Figure 1.8 plots the differences in payoffs due to self-deception for advisors and clients. For the lowest ECOST level of $0.20 each advisor loses $0.03 per decision while the client gains $0.18. Increasing ECOST shifts this relationship as advisors' self-deceptive bias is increasing as well. As a result, if ECOST equals $1.00 advisors gain $0.04 and clients lose $0.03 per decision, respectively. However, this asymmetric tendency between advisor and client payoffs is caused by the choice of the payoff matrices, which hold the differences between advisor and client payoffs constant across all ECOST levels in order to control for possible inequity aversion of the advisor. This implies that the client's potential gains are negatively related to the advisor's and are also larger in magnitude. As a consequence, the average total payoff (advisor plus client) exhibits a U-shape in ECOST and is positive for most levels of the private incentives to recommend a bad investment opportunity. Advisors' engagement in self-deception is hence welfare enhancing in this experiment. Whether this effect is solely driven by the specific payoff matrices used in this experiment or whether people engage in self-deception specifically to enhance welfare may be addressed in future research.

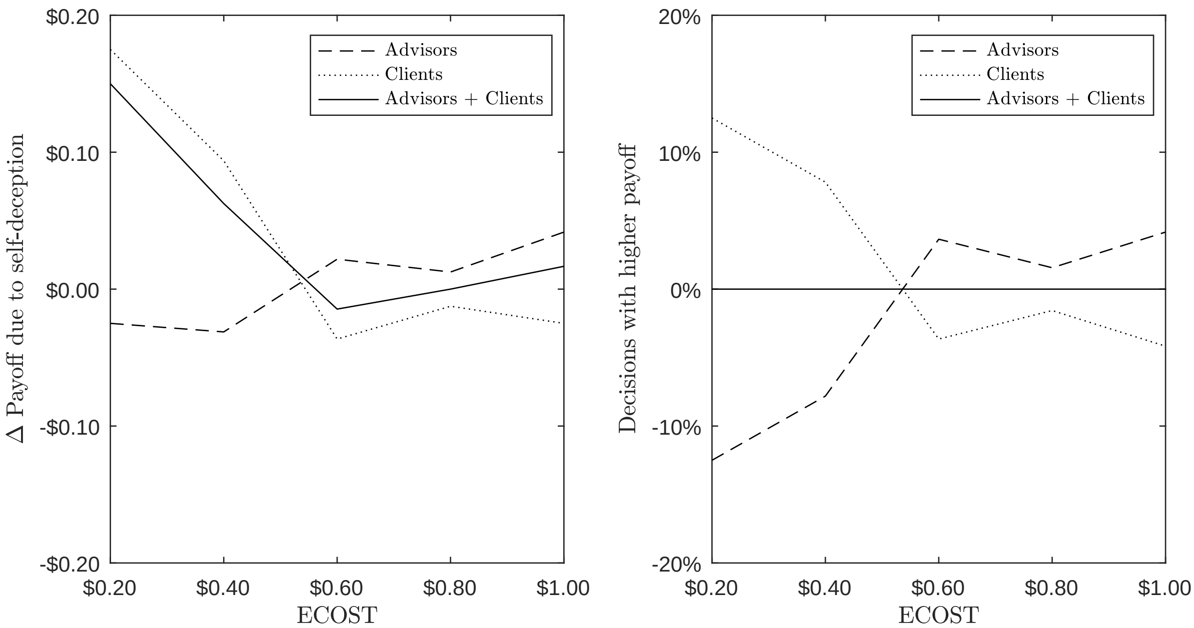


Figure 1.8: Payoff implications of self-deception in investment consulting. The left plot shows the differences in payoffs due to self-deception for advisors and clients. The right plot depicts the percentage of decisions in which a higher payoff is attained.

By contrast, considering the likelihood to be successful in achieving the high payoff, there is, by construction, a symmetric relationship between what the advisor gains and the client loses or vice versa (see right-hand side of Figure 1.8). On average, advisors (clients) end up with a lower (higher) payoff in additional 12.5% of the decisions, i.e., in every eighth decision if ECOST equals \$0.20.[38] For the highest ECOST level advisors (clients) engagement in self-deception yields a higher (lower) payoff in additional 4.2% of all decisions or differently stated in approximately every 24th decision. Since people usually make countless decisions every day such a slight increase or decrease in payoffs might accumulate to a substantial gain or loss.

[38] Since there are only two states in this setting, the claim to end up with a lower (higher) payoff is equivalent to ending up in the low (high) state.

Table 1.6: Payoff implications of self-deception in investment consulting

	Logit			OLS		
Dependent variable:	High Payoff AUT			Payoff AUT		
	(1)	(2)	(3)	(4)	(5)	(6)
ECOST	-1.158	-1.342	-2.963	39.296***	38.713***	21.601
	(1.121)	(1.143)	(2.301)	(13.444)	(13.205)	(23.549)
ECOST^2	0.774	0.840	0.869	-34.669***	-34.947***	-34.837***
	(0.884)	(0.896)	(0.908)	(11.357)	(11.167)	(11.184)
Fundamental bias	-0.008	-0.345	-0.339	-0.062	-4.156	-4.088
	(0.327)	(0.301)	(0.299)	(4.200)	(4.073)	(4.046)
$(\text{Fundamental bias})^2$		-0.744**	-0.762**		-7.616*	-7.760*
		(0.314)	(0.309)		(4.293)	(4.250)
Self-deceptive bias	0.489*	0.507*	0.526*	7.921**	7.669**	7.839**
	(0.251)	(0.270)	(0.269)	(3.282)	(3.414)	(3.384)
$(\text{Self-deceptive bias})^2$		-0.170	-0.162		-3.096	-3.006
		(0.255)	(0.255)		(3.238)	(3.254)
Sex	0.254	0.246	0.245	2.670	2.472	2.467
	(0.305)	(0.298)	(0.298)	(3.912)	(3.799)	(3.798)
Age	-0.001	-0.006	-0.006	-0.012	-0.066	-0.066
	(0.017)	(0.017)	(0.017)	(0.230)	(0.228)	(0.227)
Education	-0.065	-0.110	-0.111	-1.066	-1.577	-1.576
	(0.232)	(0.227)	(0.227)	(2.936)	(2.849)	(2.849)
Income	0.100	0.104	0.104	1.040	1.117	1.118
	(0.115)	(0.114)	(0.114)	(1.485)	(1.438)	(1.438)
Religion	-0.236	-0.185	-0.185	-2.802	-2.164	-2.146
	(0.293)	(0.289)	(0.290)	(3.709)	(3.604)	(3.608)
PV	-0.611***	-0.592***	-0.684**	-8.489***	-8.075***	-9.301***
	(0.173)	(0.174)	(0.289)	(2.054)	(2.091)	(3.516)
ECOST × PV			0.151			2.045
			(0.305)			(3.117)
Prosoc	-0.053	-0.033	-0.206	-0.880	-0.638	-2.142
	(0.093)	(0.089)	(0.166)	(1.251)	(1.138)	(2.228)
ECOST × Prosoc			0.288			2.515
			(0.187)			(2.190)
Numskill	-0.040	-0.066	-0.067	0.144	-0.231	-0.232
	(0.133)	(0.138)	(0.139)	(1.664)	(1.688)	(1.689)
Constant	3.710***	4.135***	5.112***	156.435***	160.121***	170.345***
	(1.148)	(1.152)	(1.690)	(14.654)	(14.276)	(20.153)
Observations	960	960	960	960	960	960
Number of participants	96	96	96	96	96	96
R-squared	0.084	0.099	0.101	0.109	0.122	0.123

This table presents coefficients of logit regressions in columns (1), (2) and (3), and OLS regressions in columns (4), (5) and (6). The dependent variable in columns (1), (2) and (3) is High Payoff AUT which is a binary variable that takes the value one whenever the advisor ended up in the high payoff state in AUT, and zero else. The dependent variable in columns (4), (5) and (6) is the payoff in cents in each round in AUT. ECOST denotes economic costs of stating the truth, i.e., the monetary incentives to recommend a bad investment opportunity. The fundamental bias corresponds to EX:R−EX:A and the self-deceptive bias to AUT−EX:R. Sex, Education, Income, and Religion are as defined above. PV reflects an index of protected values for honesty. Prosoc captures prosocial concerns. Numskill is the number of correctly solved numerical patterns in the logical test. Robust standard errors are obtained by clustering at the individual level and appear in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

1.5 Neutral Setting

In order to test the general validity of the findings in the investment consulting setting, I conduct a second experiment with a neutral framing.

1.5.1 Experimental Design

I slightly adapt a well-established sender-receiver game from the literature. À la Erat and Gneezy (2012), nature rolls a 6-sided die in the beginning. The better-informed party (the Sender) privately observes the outcome from the roll of the 6-sided die. In addition, the Sender is informed about the two payoff options: Option X if the uninformed party (the Receiver) chooses a number that coincides with the actual number from the roll of the die, and Option Y if the Receiver chooses any number different from the actual outcome of the roll of the 6-sided die. Knowing the payoff matrix, the Sender must send a message about the actual outcome to the Receiver. The message space consists of six messages:

1. "The outcome from the roll of the 6-sided die is 1."
2. "The outcome from the roll of the 6-sided die is 2."
3. "The outcome from the roll of the 6-sided die is 3."
4. "The outcome from the roll of the 6-sided die is 4."
5. "The outcome from the roll of the 6-sided die is 5."
6. "The outcome from the roll of the 6-sided die is 6."

The Receiver does not (and will never) know the payoffs. He chooses one number upon receiving the Sender's message. The payments to both players solely depend on whether the Receiver picked the actual outcome of the roll of the die or not.

All payoffs as well as the experimental procedure including the decision stages are similar to the investment consulting setting in order to facilitate comparability. The Sender's payoff is maximized if the Receiver chooses any other number than the actual outcome of the roll of the die, hence she has incentives to tell a wrong number and thereby lie in order to maximize her payoff.

1.5.2 Data

I recruited 300 additional US participants on mTurk. The subjects are required to have similar qualifications as in the investment consulting experiment, i.e., successfully completed at least 1000 human intelligence tasks (HITs) with an overall approval rate of more than 95%. In addition, there are two understanding questions. The first understanding question was adequately solved by 63% of all participants. 97% were able to answer the second question correctly and thereby showed that they understood the mechanisms of the game perfectly.[39]

The sample consists of 54% men and 46% women from 19 to 67 years with highest completed education from high school to graduate school. The median participant is 33 years old, holds an undergraduate degree and has a household income of $50′000 just like in the investment consulting experiment. Participants' earnings were a $0.25 show-up fee plus an average bonus of about $2.86 for advisors and $1.73 for clients. Table 1.7 summarizes the sample's descriptive statistics. In order to derive revealed first-order beliefs according to the methodology of Andreoni and Sanchez (2014), I apply similar exclusion criteria to refine the sample as in the investment consulting experiment, i.e., I drop participants who switch multiple times between preferring the lottery and the outcome of the game, and those who always prefer the lottery or the outcome of the game. In total, 37 participants are excluded. For the sake of brevity, I do neither discuss descriptive statistics of stated and revealed beliefs nor their relationship in the neutral setting but refer to Figure 1.C.3 and Figure 1.C.4 in Appendix 1.C for a detailed illustration.

1.5.3 Unethical Behavior in the Neutral Setting

Overall, dishonesty in AUT ranges from 50.4% for ECOST of $0.20 to 81.9% for ECOST of $1.00 (see Figure 1.9). This compares to a proportion of lies between 43.2% and 72.9% in the investment consulting experiment in Section 1.4. Lying is also slightly more pronounced than in Erat and Gneezy's (2012) underlying study, which observes lying rates from 43% to 76% depending on the considered payoff matrices. However, they find the lowest proportion of lies (43%) in the case of altruistic white lies, i.e., when the Sender lies (at a small cost) in order to increase the Receiver's payoff by a substantial amount. The payoff matrices in this paper do not consider white lies. It is therefore unsurprising that I find slightly more lying than Erat and Gneezy (2012). The proportion of lies is increasing in ECOST and flattening for higher incentive levels.

Table 1.8 confirms the evidence of the investment consulting setting. Lying is significantly increasing and concave in the incentives provided to lying. In addition, I find lying to be significantly decreasing in the Sender's stated first-order belief indicating that lying costs might be increasing in the first-order belief. If the first-order belief did not have an effect on lying costs, a Sender's expected payoff would be increasing in the first-

[39] All findings are robust to the exclusion of participants who did not answer the understanding questions adequately.

Table 1.7: Descriptive statistics - Neutral setting

	Mean	Std.	Min	Max	Obs.
Sex	0.54	0.50	0	1	300
Age	35.50	10.39	19	67	300
Education	2.77	0.64	2	4	300
Income	2.91	1.45	0	6	300
Religion	0.46	0.50	0	1	300
PV	4.28	1.09	1.56	7.00	300
Prosoc	3.49	1.83	1.00	7.00	300
Stated belief	0.49	0.26	0.00	1.00	150
Revealed belief	0.46	0.24	0.00	1.00	113
Numskill	1.6	1.31	0	6	300
Bonus	2.29	0.78	0.60	3.90	300

This table shows descriptive statistics for all Senders and Receivers. Sex is a binary variable taking the value one if the participant reports to be male, and zero if female. Education corresponds to the highest completed education stage and takes the following values: 0: No schooling, 1: Elementary school, 2: High school, 3: Undergraduate degree, 4: Graduate school. Income reflects yearly household income and takes the following values: 1: $< \$20'000$, 2: $\$20'000 - 39'999$, 3: $\$40'000 - 59'999$, 4: $\$60'000 - 79'999$, 5: $\$80'000 - 99'999$, 6: $> \$100'000$. Religion denotes a binary variable which takes the value one if the participant indicates to belong either to Christianity, Islam, Judaism, Hinduism, Buddhism or a Folk religion, and zero else. PV reflects an index of protected values for honesty borrowed from Tanner et al. (2009). Prosoc is an index capturing prosocial concerns. The full set of questions is depicted in Appendix 1.D. Stated belief is the advisor's self-reported first-order belief about the client's behavior. Revealed belief reflects the action-implied first-order belief according to the methodology of Andreoni and Sanchez (2014). Numskill is the number of correctly solved numerical patterns in the logical test. Bonus corresponds to the variable payoff achieved.

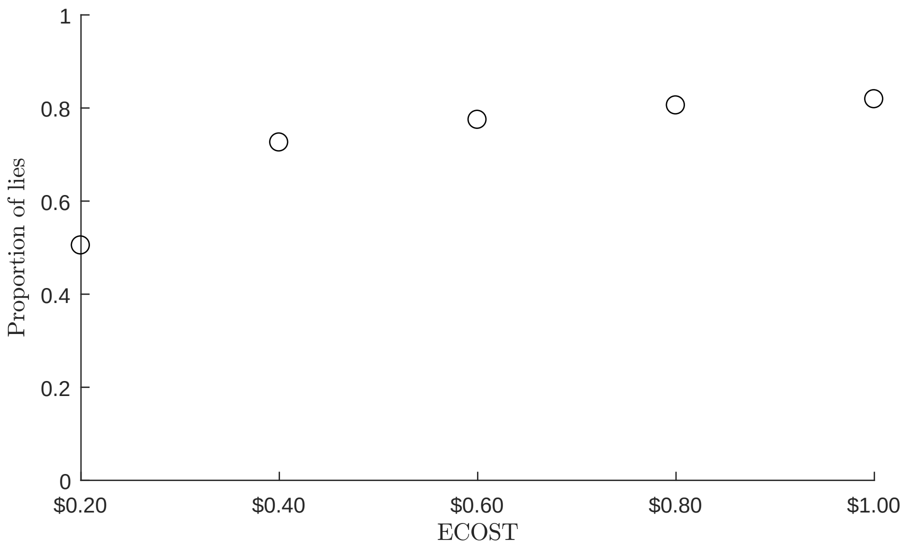


Figure 1.9: Lying in AUT across ECOST levels in the neutral setting. Each circle corresponds to the average proportion of lies in two situations with similar ECOST.

Table 1.8: Determinants of lying behavior in the neutral setting

Dependent variable:	Lie					
	(1)	(2)	(3)	(4)	(5)	(6)
ECOST	1.735***	6.271***	1.819***	6.238***	2.006***	8.162***
	(0.240)	(1.767)	(0.254)	(1.844)	(0.309)	(2.270)
ECOST^2		-2.224***		-2.349***		-3.541***
		(0.837)		(0.878)		(0.957)
Stated belief			-1.525***	-1.531***		
			(0.454)	(0.457)		
Revealed belief					0.646	0.642
					(0.557)	(0.564)
Sex	-0.251	-0.253	-0.191	-0.192	-0.302	-0.305
	(0.282)	(0.284)	(0.284)	(0.286)	(0.320)	(0.325)
Age	-0.032**	-0.032**	-0.031**	-0.031**	-0.037**	-0.038**
	(0.014)	(0.014)	(0.014)	(0.014)	(0.017)	(0.017)
Education	-0.247	-0.253	-0.241	-0.248	-0.060	-0.065
	(0.234)	(0.236)	(0.211)	(0.214)	(0.272)	(0.277)
Income	0.047	0.049	0.003	0.004	-0.019	-0.020
	(0.099)	(0.100)	(0.103)	(0.103)	(0.108)	(0.110)
Religion	-0.277	-0.276	-0.093	-0.092	-0.158	-0.153
	(0.286)	(0.288)	(0.289)	(0.290)	(0.320)	(0.325)
PV	-0.561***	-0.278	-0.552***	-0.290	-0.627***	-0.282
	(0.190)	(0.194)	(0.184)	(0.203)	(0.225)	(0.250)
ECOST × PV		-0.510*		-0.470		-0.637*
		(0.283)		(0.298)		(0.373)
Prosoc	-0.044	-0.060	-0.085	-0.115	-0.116	-0.201
	(0.105)	(0.114)	(0.106)	(0.120)	(0.107)	(0.156)
ECOST × Prosoc		0.030		0.057		0.151
		(0.157)		(0.164)		(0.198)
Numskill	0.064	0.068	0.038	0.042	0.094	0.101
	(0.092)	(0.093)	(0.090)	(0.091)	(0.103)	(0.106)
Constant	4.126***	2.486*	4.915***	3.379**	4.042***	2.082
	(1.216)	(1.331)	(1.179)	(1.377)	(1.314)	(1.615)
Observations	1,500	1,500	1,500	1,500	1,130	1,130
Number of participants	150	150	150	150	113	113
(Pseudo) R^2	0.092	0.098	0.126	0.131	0.106	0.119

This table presents coefficients of logit regressions in AUT. The dependent variable is the binary variable Lie which takes the value one if the Sender tells a lie, and zero else. ECOST denotes economic costs of stating the truth, i.e., the monetary incentives to lie. Stated belief reflects the Sender's stated first-order belief about the likelihood that the client's action will match the message received. Revealed belief denotes the Sender's first-order belief as implied by the methodology of Andreoni and Sanchez (2014). Sex is a binary variable taking the value one if the participant reports to be male, and zero if female. Education takes the following values: 0: No schooling, 1: Elementary school, 2: High school, 3: Undergraduate degree, 4: Graduate school. Income reflects yearly household Income and takes the following values: 1: $<$ \$20′000, 2: \$20′000 – 39′999, 3: \$40′000 – 59′999, 4: \$60′000 – 79′999, 5: \$80′000 – 99′999, 6: $>$ \$100′000. Religion denotes a binary variable which takes the value one if the participant indicates to belong either to Christianity, Islam, Judaism, Hinduism, Buddhism or a Folk religion, and zero else. Numskill is the number of correctly solved numerical patterns in the logical test. PV reflects an index of protected values for honesty. Prosoc captures prosocial concerns. Robust standard errors are obtained by clustering at the individual level and appear in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

order belief and hence one would also expect lying to be increasing in the belief. Another possibility is that Senders actively misstate their beliefs in order to pretend to be more ethical. As a result, Senders who lie often would state low beliefs and honest Senders high beliefs, hence yielding the negative relationship between first-order beliefs and lying. The positive (but not significant) effect of the revealed first-order belief on lying indicates that, in fact, people misstate their beliefs, leading to an inverse relation between lying and stated beliefs.

Moreover, the proportion of lies is decreasing in age suggesting that older Senders lie less often. By contrast, in the investment consulting setting older advisors do not behave unethically less often. Protected values, the interaction term with ECOST, and preferences for prosocial concerns are negatively related to lying.

1.5.4 Bias Decomposition in the Neutral Setting

In Figure 1.10, I conduct the bias decomposition based on behavior for the three relevant cases: AUT where Senders might self-deceive, EX:A where the correct following rate (89.4%) is stated,[40] and EX:R where each Sender's revealed belief is induced. The total bias equals AUT−EX:A, the fundamental bias is EX:R−EX:A, and the self-deceptive bias reflects AUT−EX:R.

Figure 1.10 shows similar but slightly more pronounced patterns as Figure 1.6 in the investment consulting case. While the total bias decreases from 25.7 to 5.3 percentage points in absolute terms, the self-deceptive bias increases from −13.7 to 12.0 percentage points being significantly different from zero at both ends. The fundamental bias stays roughly constant around minus 15 percentage points.

Table 1.9 further generalizes the findings from the investment consulting experiment. The total bias is non-linearly decreasing in ECOST because of an increasing self-deceptive bias whose effect slightly dampens for greater levels of ECOST. As predicted, the fundamental bias is not sensitive to the incentives provided to lie as the payoffs are unknown to the Receiver. Overall, the results in the neutral setting are more pronounced than in investment consulting. Presumably, the investment consulting setting is more complex to understand producing slightly more noise.

In addition, I find the self-deceptive bias to be decreasing in the degree of education and religious beliefs. Well-educated and religious participants are less prone to self-deceptive practices (if incentives are large). Measures for protected values for honesty and preferences for prosocial concerns have no significant impact. The distribution of self-deceptive

[40] Since I only have behavioral data for 21 different induced beliefs, I use the induced belief that is closest to the actual following rate. The closest available induced belief is 90%.

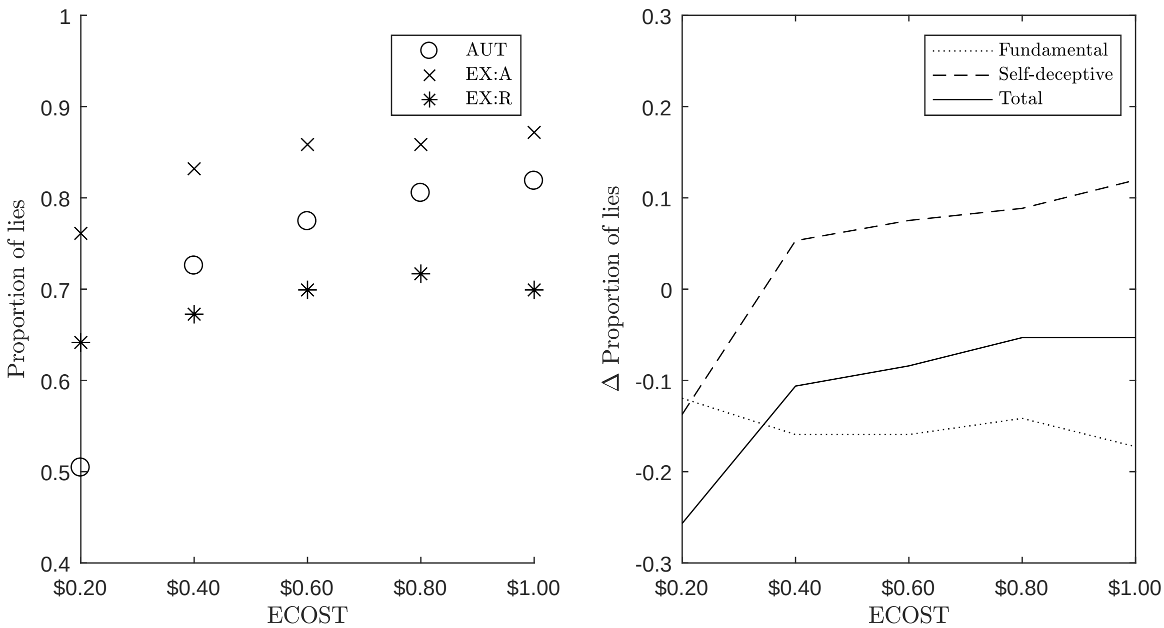


Figure 1.10: Lying in AUT, EX:A and EX:R (left), and the bias decomposition in the neutral setting (right). Circles indicate the proportion of lies for given ECOST in AUT. Crosses correspond to the proportion of lies in EX:A, and asterisks denote the proportion of lies in EX:R. The fundamental bias is computed as EX:R−EX:A. The self-deceptive bias corresponds to AUT−EX:R. The total bias equals the sum of fundamental and self-deceptive bias.

biases is illustrated in Figure 1.11. Almost 30% of the Senders have on average no self-deceptive bias.

Finally, I analyze the effect of limiting self-deception on lying by comparing behavior in AUT and EX:R. Table 1.10 provides additional evidence that the Senders engage in self-deception by distorting their first-order beliefs about the Receivers' behavior. The effect of inducing the revealed first-order belief on lying is slightly positive for small incentives but decreases sharply for higher ECOST. If incentives are moderate the effect on lying is significantly negative. As a result, self-deception leads to increased dishonesty if incentives to behave unethically are substantial.

1.5.5 Payoff Implications in the Neutral Setting

Table 1.11 analyzes the impact of self-deception on participants' success and consequently their compensation. The first dependent variable "High Payoff AUT" measures whether the Sender managed to receive the maximum payoff, i.e., whether the Sender's message led the Receiver to choose a wrong number. It takes the value one whenever the Sender is successful in maximizing her payoff, and zero else. The second dependent variable "Payoff AUT" denotes the effective payoff the Sender achieved in AUT.

Table 1.9: Determinants of the bias decomposition in the neutral setting

Dependent variable:	Total bias		Fundamental bias		Self-deceptive bias	
	(1)	(2)	(3)	(4)	(5)	(6)
ECOST	0.856***	1.060**	-0.120	-0.145	0.976***	1.204**
	(0.259)	(0.432)	(0.159)	(0.325)	(0.263)	(0.480)
ECOST^2	-0.521***	-0.521***	0.063	0.063	-0.585***	-0.585***
	(0.196)	(0.196)	(0.116)	(0.116)	(0.197)	(0.198)
Sex	0.030	0.030	0.098	0.098	-0.068	-0.068
	(0.055)	(0.055)	(0.074)	(0.074)	(0.073)	(0.073)
Age	-0.004	-0.004	-0.002	-0.002	-0.002	-0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Education	-0.023	-0.023	0.090	0.090	-0.114*	-0.114*
	(0.035)	(0.035)	(0.055)	(0.055)	(0.060)	(0.060)
Income	-0.003	-0.003	-0.000	-0.000	-0.003	-0.003
	(0.018)	(0.018)	(0.023)	(0.023)	(0.021)	(0.021)
Religion	0.010	0.010	0.150**	0.150**	-0.140**	-0.140**
	(0.058)	(0.058)	(0.072)	(0.072)	(0.069)	(0.069)
PV	0.003	0.027	0.049	0.053	-0.046	-0.025
	(0.024)	(0.048)	(0.043)	(0.050)	(0.042)	(0.053)
ECOST × PV		-0.041		-0.007		-0.034
		(0.059)		(0.045)		(0.063)
Prosoc	-0.006	0.002	-0.022	-0.031	0.016	0.034
	(0.019)	(0.028)	(0.022)	(0.027)	(0.024)	(0.034)
ECOST × Prosoc		-0.014		0.015		-0.029
		(0.034)		(0.033)		(0.046)
Numskill	0.006	0.006	-0.027	-0.027	0.033	0.033
	(0.015)	(0.015)	(0.026)	(0.026)	(0.027)	(0.027)
Constant	-0.198	-0.320	-0.450	-0.436	0.253	0.116
	(0.243)	(0.337)	(0.334)	(0.367)	(0.346)	(0.398)
Observations	1,130	1,130	1,130	1,130	1,130	1,130
Number of participants	113	113	113	113	113	113
R-squared	0.039	0.039	0.065	0.066	0.078	0.079

This table presents coefficients of OLS regressions. The dependent variables are the total bias in columns (1) and (2), the fundamental bias in columns (3) and (4), and the self-deceptive bias in columns (5) and (6). The total bias is computed as AUT−EX:A, the fundamental bias as EX:R−EX:A, and the self-deceptive bias as AUT−EX:R. ECOST denotes economic costs of stating the truth, i.e., the monetary incentives to lie. Sex is a binary variable taking the value one if the participant reports to be male, and zero if female. Education takes the following values: 0: No schooling, 1: Elementary school, 2: High school, 3: Undergraduate degree, 4: Graduate school. Income reflects yearly household income and takes the following values: 1: $< \$20'000$, 2: $\$20'000 - 39'999$, 3: $\$40'000 - 59'999$, 4: $\$60'000 - 79'999$, 5: $\$80'000 - 99'999$, 6: $> \$100'000$. Religion denotes a binary variable which takes the value one if the participant indicates to belong either to Christianity, Islam, Judaism, Hinduism, Buddhism or a Folk religion, and zero else. PV reflects an index of protected values for honesty. Prosoc captures prosocial concerns. Numskill is the number of correctly solved numerical patterns in the logical test. Robust standard errors are obtained by clustering at the individual level and appear in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

Table 1.11 provides evidence that a Sender's payoff is significantly increasing in her self-deceptive bias. By contrast, a large fundamental bias in absolute terms leads to a decrease in success since it is negative for most Senders and incentive levels. I also find that people who lie less often, i.e., older and those with strong protected values for honesty,

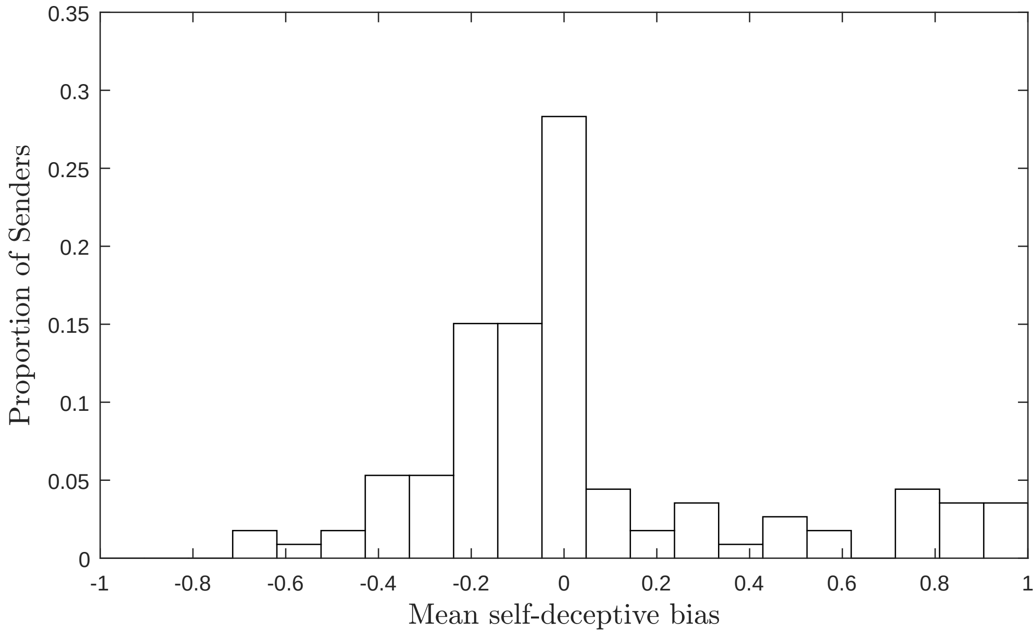


Figure 1.11: Cross-section of self-deceptive biases in the neutral setting. The self-deceptive bias is computed as the difference between behavior in AUT and EX:R averaged over all ECOST levels.

Table 1.10: Limiting self-deception in the neutral setting

Dependent variable:		Lie	
	(1)	(2)	(3)
ECOST	3.765***	4.411***	4.943***
	(0.708)	(0.727)	(1.320)
ECOST^2	-2.253***	-2.139***	-2.079***
	(0.535)	(0.543)	(0.540)
EX:R	-0.205	0.630***	0.634***
	(0.183)	(0.227)	(0.225)
ECOST $\times$ EX:R		-1.471***	-1.487***
		(0.306)	(0.307)
Controls	Yes	Yes	Yes
Interactions	No	No	Yes
Observations	2,260	2,260	2,260
Number of participants	113	113	113
(Pseudo) R^2	0.054	0.061	0.064

This table presents coefficients of logit regressions. The dependent variable is the binary variable Lie which takes the value one if the Sender tells a lie, and zero else. ECOST denotes economic costs of stating the truth, i.e., the monetary incentives to lie. EX:R is a binary variable variable which takes the value one in EX:R where every Sender's revealed belief is induced. Controls include Sex, Age, Education, Income, Religion, PV, Prosoc, and Numskill. Interactions indicates interaction terms of PV and Prosoc with ECOST. Robust standard errors are obtained by clustering at the individual level and appear in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

achieve lower payoffs due to the characteristics of the payoff matrices. In order to provide incentives to behave unethically, the expected return of lying is positive as long as the first-order belief is greater than 1/6.

Next, I match Sender behavior in EX:R to Receiver behavior in AUT and derive the resulting payoffs. As a consequence, I can compare the payoffs in AUT to the payoffs they would have achieved if they played without any self-deceptive bias, i.e., as in EX:R. Figure 1.12 shows the differences in payoffs between AUT and EX:R, and the effect of self-deception on the compensation of Senders and Receivers.

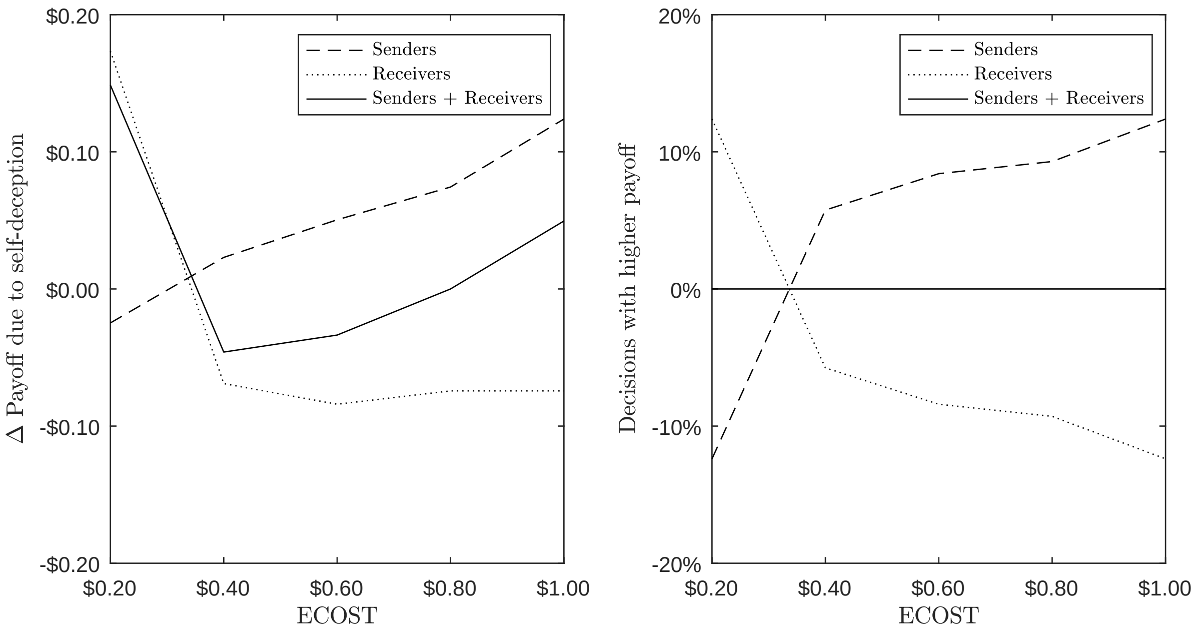


Figure 1.12: Payoff implications of self-deception in the neutral setting. The left plot shows the differences in payoffs due to self-deception for Senders and Receivers. The right plot depicts the percentage of decisions in which a higher payoff is attained.

Self-deception decreases the Sender's payoff by \$0.03 per decision if ECOST is \$0.20 and increases it by \$0.12 if ECOST equals \$1.00. In contrast, the Receiver gains \$0.17 per decision for the lowest ECOST level and loses \$0.07 for the highest incentive level. Combined, self-deception increases the total surplus for very low and very high incentives to lie. Analogously, by engaging in self-deception the Sender ends up in the best state in additional 12.4% of all decisions with ECOST of \$1.00, i.e., in every eight decision.

Since all results regarding biases and engagement in self-deception are consistent with the investment advisory setting, the findings in the neutral setting suggest that all results may also be valid in more general situations where lying takes place and not only in the specific case of investment advisory. Thanks to the general structure of the experiments, this paper's findings may also transfer to many other applications with a similar environment such as dishonest communication of CEOs, physicians recommending specific medication or a salesperson selling credence goods.

Table 1.11: Payoff implications of self-deception in the neutral setting

	Logit			OLS		
Dependent variable:	High Payoff AUT			Payoff AUT		
	(1)	(2)	(3)	(4)	(5)	(6)
ECOST	-3.186***	-3.289***	-3.774**	36.067***	35.930***	33.933**
	(1.215)	(1.247)	(1.848)	(11.872)	(11.940)	(16.478)
$ECOST^2$	2.152**	2.218**	2.230**	-34.379***	-34.312***	-34.348***
	(0.962)	(0.983)	(0.985)	(9.820)	(9.848)	(9.856)
Fundamental bias	0.645**	0.508*	0.512*	7.355**	6.630**	6.684**
	(0.305)	(0.282)	(0.282)	(3.352)	(3.249)	(3.244)
$(\text{Fundamental bias})^2$		-0.402	-0.406		-3.020	-3.043
		(0.268)	(0.267)		(2.823)	(2.816)
Self-deceptive bias	0.696***	0.790***	0.795***	8.484***	9.220***	9.231***
	(0.242)	(0.281)	(0.281)	(2.814)	(3.037)	(3.032)
$(\text{Self-deceptive bias})^2$		0.096	0.104		1.330	1.435
		(0.257)	(0.258)		(2.727)	(2.738)
Sex	-0.220	-0.206	-0.205	-1.922	-1.825	-1.827
	(0.282)	(0.281)	(0.281)	(3.178)	(3.173)	(3.175)
Age	-0.033**	-0.032**	-0.032**	-0.335*	-0.329*	-0.329*
	(0.014)	(0.014)	(0.014)	(0.178)	(0.179)	(0.180)
Education	-0.065	-0.064	-0.064	-0.943	-0.937	-0.940
	(0.253)	(0.258)	(0.258)	(2.753)	(2.765)	(2.767)
Income	-0.038	-0.034	-0.033	-0.512	-0.459	-0.458
	(0.096)	(0.094)	(0.094)	(1.082)	(1.081)	(1.082)
Religion	-0.240	-0.218	-0.217	-2.299	-2.133	-2.138
	(0.288)	(0.285)	(0.285)	(3.267)	(3.235)	(3.238)
PV	-0.547***	-0.563***	-0.656**	-6.460**	-6.503**	-7.372**
	(0.210)	(0.211)	(0.292)	(2.581)	(2.569)	(3.432)
ECOST × PV			0.151			1.455
			(0.235)			(2.304)
Prosoc	-0.043	-0.047	-0.022	-0.178	-0.188	0.446
	(0.094)	(0.092)	(0.137)	(1.007)	(0.990)	(1.349)
ECOST × Prosoc			-0.041			-1.057
			(0.152)			(1.284)
Numskill	0.101	0.097	0.097	1.193	1.156	1.157
	(0.099)	(0.099)	(0.099)	(0.957)	(0.958)	(0.959)
Constant	5.925***	6.017***	6.307***	167.313***	167.277***	168.478***
	(1.184)	(1.193)	(1.499)	(13.634)	(13.612)	(16.250)
Observations	1,130	1,130	1,130	1,130	1,130	1,130
Number of participants	113	113	113	113	113	113
R-squared	0.073	0.076	0.076	0.077	0.078	0.079

This table presents coefficients of logit regressions in columns (1), (2) and (3), and OLS regressions in columns (4), (5) and (6). The dependent variable in columns (1), (2) and (3) is High Payoff AUT which is a binary variable that takes the value one whenever the Sender ended up in the high payoff state in AUT, and zero else. The dependent variable in columns (4), (5) and (6) is the payoff in cents in each round from AUT. The fundamental bias corresponds to EX:R−EX:A and the self-deceptive bias to AUT−EX:R. ECOST denotes economic costs of stating the truth, i.e., the monetary incentives to lie. Sex, Education, Income, and Religion are as defined above. PV reflects an index of protected values for honesty. Prosoc captures prosocial concerns. Numskill is the number of correctly solved numerical patterns in the logical test. Robust standard errors are obtained by clustering at the individual level and appear in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

1.6 Conclusion

This paper studies self-deception to promote unethical advice by distorting one's own belief about how other people behave. My findings provide evidence that people's beliefs are biased and that these biases influence unethical behavior. In order to investigate whether people's biases are caused by the general inability to assess others' behavior per se or by people self-deceiving themselves by using motivated beliefs, I construct an experimental methodology to decompose this bias into two components: a fundamental (incentive-independent) bias, which is caused by lack of knowledge, and a self-deceptive bias, which is caused by motivated beliefs and might depend on the magnitude of private monetary benefits associated with unethical behavior.

I find that people engage in self-deception and show the self-deceptive bias to be increasing in the private incentives provided to behave unethically. By engaging in self-deception, the deceiver gains money in every decision if the monetary incentives to lie are high and loses if the incentives are small. Being fundamentally biased leads to financial losses regardless of the direction of the bias. Setting up an environment where self-deception is limited might thus promote truth-telling among individuals. Whether releasing descriptive information about behavior per se reduces lying also depends on the average fundamental bias of the population. This paper suggests that in an environment where the fundamental bias is comparably small and the incentives to lie are relatively high, simply providing descriptive behavioral information to the informed party is enough to cost-efficiently foster ethical behavior.

This study's findings hold in a general strategic information transmission setting as well as in a specific application to investment consulting and hence suggest that self-deceptive biases are important in a broad range of environments where lying and more generally unethical behavior takes place. Although I find self-deception to have a significant and economically relevant effect on truth-telling, this experimental design does not allow to quantify the degree to which individuals bias their beliefs as this needed further assumptions on the structural form of people's moral costs. I leave the quantification of the adjustments in individual beliefs to further research.

Appendices

1.A Appendix A: A Model of Self-deception

My model builds on a simplified setting of Crawford and Sobel's (1982) strategic information transmission framework. I assume that after the better-informed player (the advisor or more generally, and henceforth, the Sender) privately learns the state of the world, she is allowed to send a message about the state to the uninformed player (the client or more generally the Receiver) before the latter takes an action which determines the payoff of both players. The Sender is said to lie if her statement does not match her private information. In the investment consulting experiment this corresponds to an advisor not recommending the investment, which she knows to dominate the other investment opportunities. People supposedly trade off the monetary benefits of lying against the moral costs of unethical behavior when making their decisions. I propose an extended utility specification capturing moral costs, which can be separated into two terms: First, the cost of behaving unethically per se and second, a term that derives from the negative consequences imposed on others. The structure allows for many popular specifications in the literature.

Intuitively, the higher the probability that the Receiver will believe the Sender's message and choose the corresponding action, the more painful it is for a Sender to tell a lie. The Sender's estimate of this probability is her first-order belief. Due to the subjectivity of this first-order belief, a Sender may engage in self-deception by under- or overestimating the respective probability. The model allows for such self-deceptive strategies while not drastically restricting the composition of moral costs.[41]

1.A.1 Strategic Information Transmission

I focus on strategic information transmission games with two players. Each game features a 2-by-2 payoff matrix only known to the Sender. The information structure of the game (in contrast to the payoff structure) is common knowledge. Let Θ, $\mathcal{M}$, and $\mathcal{A}$ be discrete and finite sets. The Sender privately observes a realization $\theta \in \Theta$. Then, the Sender is asked to transmit a single message $m \in \mathcal{M}$ to the Receiver. By definition, any message in $\mathcal{M}$ is non-binding and non-verifiable. Based on this message, the Receiver takes an action $a \in \mathcal{A}$ which determines the welfare of all parties. The mapping $\sigma : \Theta \to \mathcal{M}$ is a

[41] The model is related to Rabin (1995) who studies self-serving biases in the context of fairness concerns and argues that an individual may engage in costly self-deception in order to legitimize selfish behavior. However, the model does not feature strategic information transmission between players.

strategy for the Sender and $\alpha : \mathcal{M} \rightarrow \mathcal{A}$ for the Receiver. Let the Sender enjoy utility

$$U\left(a, \theta, m, \omega, \beta\right) = \Pi^S(a, \theta) - \Psi\left(a, \theta, m, \omega, \beta\right) \mathbb{1}_{m \neq \theta},$$

where Π^S denotes the Sender's payoff function in a given state. For convenience utility is assumed to be continuous and twice-differentiable in its arguments. I posit a simple linear and additively separable utility function. The term $\Psi(\cdot)\mathbb{1}_{m \neq \theta}$ corresponds to the total moral costs a Sender incurs if she reports dishonestly. The indicator function is equal to one if message m does not coincide with the state θ (i.e., if she tells a lie), and zero else. Note that $\Psi(\cdot)$ does not necessarily depend on m and might also depend on additional parameters.

1.A.2 Moral Costs

I consider the total moral or lying costs $\Psi(\cdot)$ to be additively separable in two terms: (i) the cost of behaving unethically per se and (ii) the cost of misleading other players. While (i), once a player has chosen to lie, only depends on the structural form of the moral costs, (ii) adds an additional cost, which depends on the subjective probability that the message will indeed mislead the other player and thus impose pain on others.[42]

I postulate the following structural form

$$\Psi(a, \theta, m, \omega, \beta) = \omega \underbrace{\psi\left(a, \theta, m\right)}_{(i)} + (1 - \omega) \underbrace{f\left(\beta(\alpha(m) = m), \psi\left(a, \theta, m\right)\right)}_{(ii)}, \tag{1.A.1}$$

where $\psi(\cdot)$ denotes individual-specific moral costs and $f(\cdot)$ is a function of the first-order belief and the individual-specific moral costs with $\partial f(\cdot)/\partial \beta > 0$. $\omega \in [0, 1]$ denotes the weight she puts on term (i) relative to (ii), and $\beta(\alpha(m) = m) \in [0, 1]$ the first-order belief the Sender assigns to the Receiver's strategy $\alpha(m) = m$, i.e., the probability the Sender expects the Receiver to be credulous and to choose $a = m$, conditional on the reception of m. I simplify $\beta := \beta(\alpha(m) = m)$. If $\omega = 1$, Eq. (1.A.1) reduces to any given moral cost specification in the literature.

I do not assume a specific form of the individual moral costs $\psi(\cdot)$.[43] This setting allows for a variety of possible preferences that come up in the literature. Abeler et al. (2016) provide an extensive meta-study of existing explanations of lying. Most of these motivations consider lying costs, reputation, social norms or a combination of those. Based

[42] The cost of misleading the Receiver depends on the subjective probability rather than the Receiver's actual behavior, i.e., if he is indeed misled, as the Sender does not know whether the Receiver is going to match her message.

[43] Depending on the specification of the moral costs, several parameters may determine the structural form. For notational simplicity, I abstract from including an exhaustive list of possible arguments.

on an econometric analysis Bögli et al. (2017) find a combination of the same motivations and reference-dependent preferences to explain lying behavior best in-sample and also out-of-sample.

1.A.3 Self-deception

If there is no information about the behavior of the uninformed party, i.e., as long as the Sender forms her belief autonomously, she might engage in self-deception by under- or overestimating the likelihood that her behavior will mislead people. Due to the first-order belief's subjectivity the Sender may engage in self-deception by taking a misstated belief $\beta = \delta\hat{\beta}$ for her utility computation instead. Parameter $\delta \in \mathbb{R}^+$ reflects the degree of self-deception the Sender imposes on her belief of misleading the Receiver and might depend on several variables such as the incentives to lie or other characteristics of the decision environment. I assume people to maximize utility conditional on being self-deceptively biased rather than to maximize utility with respect to their degree of self-deception. Obviously, $\delta < 1$ ($\delta > 1$) reflects an undervaluation (overvaluation) of the subjective first-order belief $\hat{\beta}$.

If $\delta < 1$, the intuition is that the Sender underestimates the probability that the Receiver will be misled by her message in order to reduce the expected pain her behavior imposes on the Receiver. This will have two opposing effects: On the one hand, the expected monetary payoff of lying per se will decrease due to the lower chance that the Receiver chooses what the Sender recommends. On the other hand, moral costs will be lower, as the expected harm imposed on the Receiver will be smaller. Similarly, if $\delta > 1$ the expected monetary payoff will be higher but also the moral costs. As a result, the net effect of self-deception depends on the characteristics of a given game and a player's individual preferences.

As long as there is uncertainty about the actual following rate as in AUT where the Sender does not have any exogenous information about the actual strategy of the other player, she may incorporate this uncertainty by self-deceiving and distorting her (subjective) belief in order to assess the total lying costs. In case beliefs are not assessed subjectively but provided exogenously, i.e., in EX where the Receiver's strategy is known to the Sender, self-deception is not supposed to take place to the same extent due to the lack of uncertainty about the correct belief and consequently $\delta \approx 1$.[44] Table 1.A.1

[44] The parameter δ might not be exactly equal to one if first-order beliefs are provided exogenously as people could still engage in self-deception to some degree. However, I assume the room for self-deception to be limited once exogenous first-order beliefs are in place. This allows me to assume without loss of generality that $\delta = 1$. Note that, for the purpose of this paper, it would be sufficient to assume that $|\delta - 1|$ is smaller once uncertainty is released, i.e., the parameter of self-deception is closer to one if beliefs are provided exogenously.

summarizes the belief formation in AUT, EX:A, and EX:R.[45]

Table 1.A.1: Belief formation

	Self-deception δ	Elicited/Induced belief $\hat{\beta}$	Effective belief β
AUT	$\delta \in \mathbb{R}^+$	$\hat{\beta} = \hat{\beta}^{rev}$	$\beta^{AUT} = \delta\hat{\beta}^{rev}$
EX:A	$\delta = 1$	$\hat{\beta} = q$	$\beta^{EX:A} = q$
EX:R	$\delta = 1$	$\hat{\beta} = \hat{\beta}^{rev}$	$\beta^{EX:R} = \hat{\beta}^{rev}$

This table presents the belief formation in AUT, EX:A, and EX:R. δ denotes the degree of self-deception whereas $\delta = 1$ indicates that no self-deception takes place. $\hat{\beta}$ corresponds to the Sender's elicited or induced first-order belief, i.e., the revealed belief in AUT and the induced beliefs in EX. β is the Sender's effective belief used to derive expected utilities. In AUT, it corresponds to the motivated belief, i.e., the distorted elicited belief. $\hat{\beta}^{rev}$ denotes the Sender's revealed belief and q equals the Receivers' actual following rate, i.e., the probability that the Receiver will be credulous.

Having the effective first-order belief, I now define the bias in beliefs and decompose it into a fundamental and a self-deceptive component. Let the total bias be

$$\Delta_{\text{tot}} \equiv \beta^{AUT} - \beta^{EX:A} = \delta\hat{\beta}^{rev} - q, \tag{1.A.2}$$

where $\hat{\beta}^{rev}$ denotes the Sender's revealed first-order belief and q the Receivers' average following rate, i.e., the probability that the Receiver chooses the action recommended in the Sender's message. The fundamental bias then corresponds to

$$\Delta_{\text{fun}} \equiv \beta^{EX:R} - \beta^{EX:A} = \hat{\beta}^{rev} - q, \tag{1.A.3}$$

i.e., the difference between the Sender's revealed belief and the correct following rate. Finally, the self-deceptive bias is computed as

$$\Delta_{\text{sd}} \equiv \beta^{AUT} - \beta^{EX:R} = (\delta - 1)\hat{\beta}^{rev}. \tag{1.A.4}$$

Note that the bias in beliefs and its effect on unethical behavior must not necessarily be parallel. A biased belief enters the expected utility maximization twice: it influences the expected monetary payoff and it affects moral costs. Depending on the structural form of the Sender's moral costs, the impact of a biased belief on unethical behavior is determined. I illustrate a possible bias decomposition where the bias in beliefs and its effect on unethical behavior is not parallel in Figure 1.A.1. Therefore, I assume the second partial derivative of $f(\cdot)$ with respect to the first-order belief to be negative, i.e., $\partial^2 f(\cdot)/\partial^2\beta < 0$. Given the characteristics of $f(\cdot)$ the proportion of lies is then U-shaped

[45] Belief formation EX:S happens analogously to EX:R but with the Sender's stated first-order belief instead.

in her first-order belief. As a result, the direction of the bias in beliefs is not similar to its effect on lying if $\beta < \beta^{\star}$, where $\beta^{\star}$ denotes the first-order belief that minimizes the proportion of lies.

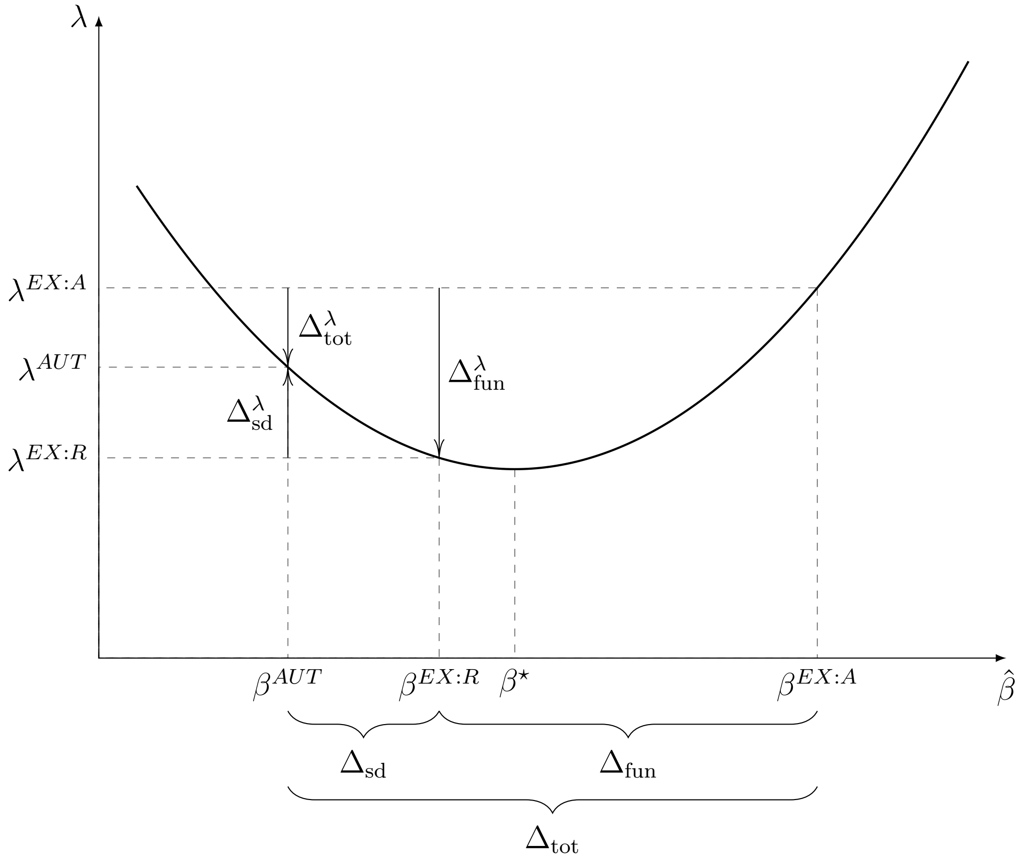


Figure 1.A.1: Decomposition of the bias in beliefs. The graph assigns a Sender's effective first-order belief β to the corresponding proportion of lies λ. Δ_{tot}, Δ_{fun}, and Δ_{sd} denote the total, the fundamental, and the self-deceptive bias in beliefs, respectively. $\Delta^{\lambda}_{\text{tot}}$, $\Delta^{\lambda}_{\text{fun}}$, and $\Delta^{\lambda}_{\text{sd}}$ correspond to the effect of the total, fundamental, and self-deceptive bias in beliefs on the proportion of lies. The total bias is defined in Eq. (1.A.2), the fundamental bias in Eq. (1.A.3), and the self-deceptive bias in Eq. (1.A.4).

In Figure 1.A.1, the Sender underestimates her first-order belief, i.e., $\delta < 1$. As β^{AUT} and $\beta^{EX:R}$ are both smaller than $\beta^{\star}$, $\beta^{AUT} < \beta^{EX:R}$ implies $\lambda^{AUT} > \lambda^{EX:R}$ where λ^{AUT} ($\lambda^{EX:R}$) denotes the proportion of lies in AUT (EX:R). Consequently, self-deception leads to more lying in AUT compared to EX:R. The resulting self-deceptive bias is positive. However, other assumptions on $f(\cdot)$ will produce different results.

1.A.4 Utility Maximization

I assume players to maximize expected utility with respect to their strategies. The Sender's conditional expected utility under the subjective probability measure β is given by

$$\mathbb{E}_\beta\Big[U(a,\theta,m,\omega,\beta)\Big|m\Big] = \mathbb{E}_\beta\left[\Pi^S(a,\theta) - \Psi(a,\theta,m,\omega,\beta)\mathbb{1}_{m\neq\theta}\Big|\, m\right],$$

which equals

$$\begin{aligned}\mathbb{E}_\beta\Big[U(a,\theta,m,\omega,\beta)\Big|m\neq\theta\Big] &= \beta\Pi^S(a\neq\theta,\theta) + (1-\beta)\mathbb{E}\left[\Pi^S(a\neq m,\theta)\Big|\, m\neq\theta\right] \\ &\quad - \Big(\omega\psi(a,\theta,m) + (1-\omega)f\left(\beta,\psi(a,\theta,m)\right)\Big),\end{aligned}$$

if she tells a lie, and

$$\mathbb{E}_\beta\Big[U(a,\theta,m,\omega,\beta)\Big|m=\theta\Big] = \beta\Pi^S(a=\theta,\theta) + (1-\beta)\mathbb{E}\left[\Pi^S(a\neq m,\theta)\Big|\, m=\theta\right],$$

if reporting truthfully.

When considering to lie people supposedly trade-off the economic cost of stating the truth (Ξ or ECOST), or equivalently the marginal benefit of lying against the individual moral costs of unethical behavior. I compute the Sender's expected ECOST to equal

$$\begin{aligned}\Xi(\beta) :=& \beta\underbrace{\left(\Pi^S(a\neq\theta,\theta) - \Pi^S(a=\theta,\theta)\right)}_{=:\tilde{\Xi}} \\ &+(1-\beta)\Big(\mathbb{E}\left[\Pi^S(a\neq m,\theta)|m\neq\theta\right] - \mathbb{E}\left[\Pi^S(a\neq m,\theta)|m=\theta\right]\Big),\end{aligned} \tag{1.A.5}$$

where $\tilde{\Xi}$ is constant and equals $\Xi(\beta=1)$, i.e., expected ECOST if $\beta=1$. Eq. (1.A.5) simplifies to

$$\Xi(\beta) = \tilde{\Xi}\Big(1-(1-\beta)\Big(\underbrace{\mathbb{P}\left(a=\theta|a\neq m,\ m\neq\theta\right)}_{=:\mathbb{P}_1} + \underbrace{\mathbb{P}\left(a\neq\theta|a\neq m,\ m=\theta\right)}_{=:\mathbb{P}_2}\Big)\Big), \tag{1.A.6}$$

where $\mathbb{P}_1$ ($\mathbb{P}_2$) is the conditional probability that the Receiver chooses $a=\theta$ ($a\neq\theta$) given that the Receiver does not follow and the Sender transmits message $m\neq\theta$ ($m=\theta$), respectively. These two probabilities depend on the characteristics of a specific game.[46]

The likelihood that the Sender tells a lie depends on the difference in expected utility between lying and telling the truth as given by

$$\mathbb{E}_\beta\Big[U(a,\theta,m,\omega,\beta)\Big|m\neq\theta\Big] - \mathbb{E}_\beta\Big[U(a,\theta,m,\omega,\beta)\Big|m=\theta\Big],$$

[46] For example, the game considered in this paper features probabilities $\mathbb{P}_1 = 1/5$ and $\mathbb{P}_2 = 1$ as there exist six messages and also six possible actions implying that even if the Receiver does not follow and the Sender transmits a message not corresponding to the true state, the Receiver's action will coincide with the true state with probability one fifth only. It follows directly from Eq. (1.A.6) that lying is expected to be less pronounced in games with high $\mathbb{P}_1 + \mathbb{P}_2$ as this reduces the incentives to lie. This is consistent with the findings of Gneezy (2005) whose game features $\mathbb{P}_1 = \mathbb{P}_2 = 1$.

which is equal to

$$\Xi(\beta) - \Psi\left(a, \theta, m, \omega, \beta\right),$$

and can be rewritten as

$$\tilde{\tilde{\Xi}}\left(1 - (1-\beta)\left(\mathbb{P}_1 + \mathbb{P}_2\right)\right) - \omega\psi(a,\theta,m) - (1-\omega)\, f\left(\beta, \psi(a,\theta,m)\right).$$

The greater this expression, the more often will the Sender lie. As a consequence, the sensitivity of the decision to lie with respect to the degree of self-deception is given by

$$\frac{\partial}{\partial\delta} = \tilde{\tilde{\Xi}}\hat{\beta}\left(\mathbb{P}_1 + \mathbb{P}_2\right) - (1-\omega)\,\frac{\partial f\left(\cdot\right)}{\partial\delta},$$

which is positive if and only if

$$\tilde{\tilde{\Xi}}\hat{\beta}\left(\mathbb{P}_1 + \mathbb{P}_2\right) > (1-\omega)\,\frac{\partial f\left(\cdot\right)}{\partial\delta}.$$

Intuitively, for individuals with $\omega = 1$, a higher degree of self-deception will increase lying, as it only influences the expected monetary payoff. However, if $\omega < 1$ the structural forms of $f(\cdot)$ and $\psi(\cdot)$ need to be known in order to determine the marginal effect of an increase in δ.

1.B Appendix B: Robustness: Stated First-order Beliefs

In order to test the robustness of the findings, I also run the whole analysis for stated instead of revealed beliefs. Taking stated beliefs as the true beliefs ignores possible strategic misstatement of beliefs.

1.B.1 Investment Consulting

1.B.1.1 Bias Decomposition

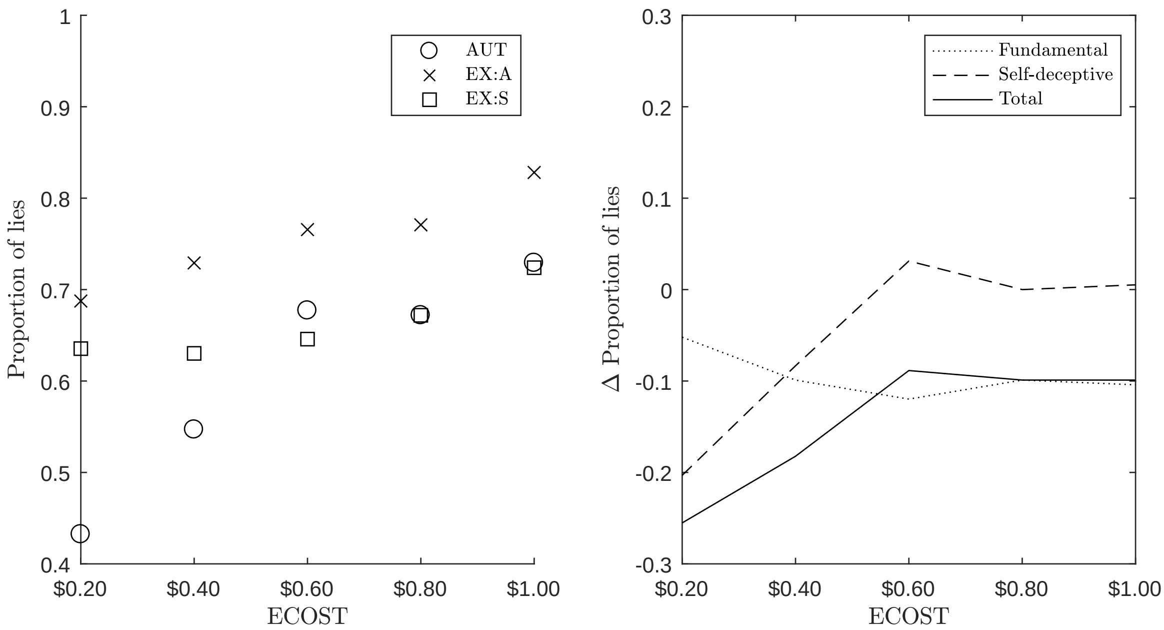


Figure 1.B.1: Unethical behavior in AUT, EX:A, and EX:S (left) and the bias decomposition in investment consulting based on stated beliefs (right). Circles indicate the proportion of lies for given ECOST in AUT. Crosses correspond to the proportion of lies in EX:A, and squares denote the proportion of lies in EX:S. The fundamental bias is computed as EX:S−EX:A. The self-deceptive bias corresponds to AUT−EX:S. The total bias equals the sum of fundamental and self-deceptive bias.

Table 1.B.1: Determinants of the biases in investment consulting based on stated beliefs

Dependent variable:	Total bias		Fundamental bias		Self-deceptive bias	
	(1)	(2)	(3)	(4)	(5)	(6)
ECOST	0.734***	1.208***	-0.320*	-0.552*	1.054***	1.760***
	(0.263)	(0.413)	(0.179)	(0.314)	(0.268)	(0.461)
$ECOST^2$	-0.446**	-0.446**	0.223*	0.223*	-0.670***	-0.670***
	(0.197)	(0.197)	(0.128)	(0.128)	(0.191)	(0.191)
Sex	0.007	0.007	0.114**	0.114**	-0.107	-0.107
	(0.053)	(0.053)	(0.055)	(0.056)	(0.069)	(0.069)
Age	0.002	0.002	0.002	0.002	0.000	0.000
	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
Education	-0.048	-0.048	-0.016	-0.016	-0.032	-0.032
	(0.037)	(0.037)	(0.053)	(0.053)	(0.057)	(0.057)
Income	0.018	0.018	0.002	0.002	0.016	0.016
	(0.019)	(0.019)	(0.014)	(0.014)	(0.021)	(0.021)
Religion	-0.123**	-0.123**	0.019	0.019	-0.142*	-0.142*
	(0.056)	(0.056)	(0.060)	(0.060)	(0.072)	(0.072)
PV	0.039	0.080*	0.050**	0.030	-0.011	0.050
	(0.026)	(0.043)	(0.023)	(0.027)	(0.031)	(0.042)
ECOST × PV		-0.067		0.033		-0.101*
		(0.048)		(0.028)		(0.056)
Prosoc	0.006	0.040	0.005	-0.011	0.000	0.051*
	(0.014)	(0.028)	(0.018)	(0.017)	(0.021)	(0.030)
ECOST × Prosoc		-0.057*		0.028		-0.084**
		(0.033)		(0.019)		(0.038)
Numskill	-0.026	-0.026	0.037*	0.037*	-0.062**	-0.062**
	(0.023)	(0.024)	(0.020)	(0.020)	(0.029)	(0.029)
Constant	-0.484**	-0.769**	-0.373*	-0.234	-0.111	-0.535*
	(0.202)	(0.299)	(0.204)	(0.229)	(0.226)	(0.287)
Observations	960	960	960	960	960	960
Number of participants	96	96	96	96	96	96
R-squared	0.051	0.054	0.067	0.068	0.063	0.069

This table presents coefficients of OLS regressions. The dependent variables are the total bias in columns (1) and (2), the fundamental bias in columns (3) and (4), and the self-deceptive bias in columns (5) and (6). The total bias is computed as AUT−EX:A, the fundamental bias as EX:S−EX:A, and the self-deceptive bias as AUT−EX:S. ECOST denotes economic costs of stating the truth, i.e., the monetary incentives to recommend a bad investment opportunity. Sex is a binary variable taking the value one if the participant reports to be male, and zero if female. Education takes the following values: 0: No schooling, 1: Elementary school, 2: High school, 3: Undergraduate degree, 4: Graduate school. Income reflects yearly household income and takes the following values: 1: $<$ \$20′000, 2: \$20′000 − 39′999, 3: \$40′000 − 59′999, 4: \$60′000 − 79′999, 5: \$80′000 − 99′999, 6: $>$ \$100′000. Religion denotes a binary variable which takes the value one if the participant indicates to belong either to Christianity, Islam, Judaism, Hinduism, Buddhism or a Folk religion, and zero else. PV reflects an index of protected values for honesty. Prosoc captures prosocial concerns. Numskill is the number of correctly solved numerical patterns in the logical test. Robust standard errors are obtained by clustering at the individual level and appear in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

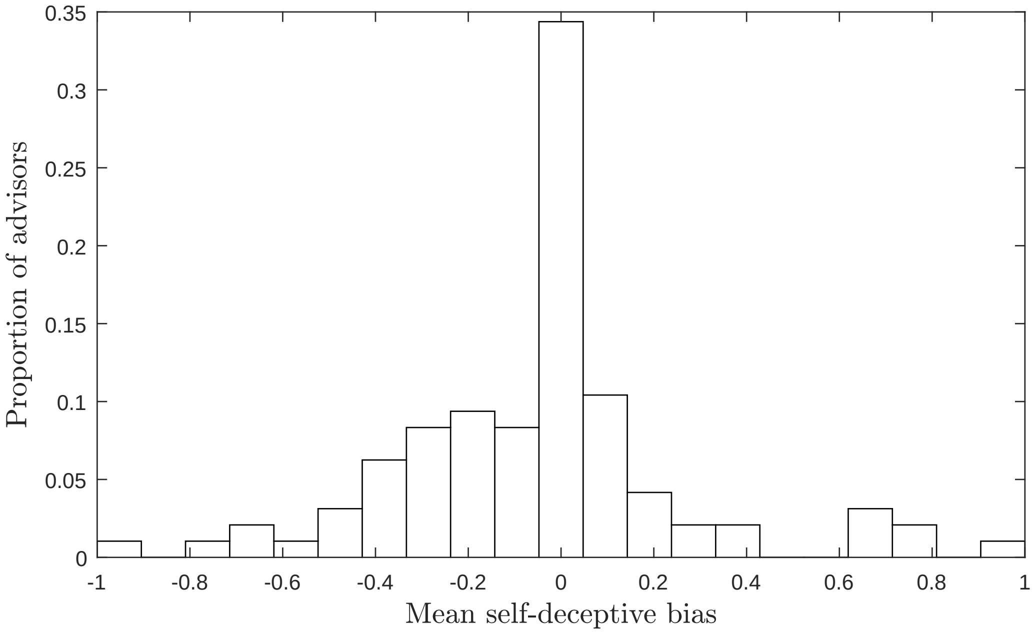


Figure 1.B.2: Cross-section of self-deceptive biases in investment consulting based on stated beliefs. The self-deceptive bias is computed as the difference between behavior in AUT and EX:S averaged over all ECOST levels.

Table 1.B.2: Limiting self-deception in investment consulting based on stated beliefs

Dependent variable:		Lie	
	(1)	(2)	(3)
ECOST	1.633**	2.213***	3.623*
	(0.780)	(0.841)	(2.012)
ECOST^2	-0.401	-0.390	-0.301
	(0.594)	(0.593)	(0.578)
EX:S	0.248	0.943***	0.928***
	(0.167)	(0.216)	(0.211)
ECOST × EX:S		-1.197***	-1.183***
		(0.311)	(0.316)
Controls	Yes	Yes	Yes
Interactions	No	No	Yes
Observations	1,920	1,920	1,920
Number of participants	96	96	96
(Pseudo) R^2	0.103	0.108	0.109

This table presents coefficients of logit regressions. The dependent variable is the binary variable Lie which takes the value one if the advisor behaves unethically, and zero else. ECOST denotes economic costs of stating the truth, i.e., the monetary incentives to recommend a bad investment opportunity. EX:S is a binary variable variable which takes the value one in EX:S where every advisor's stated belief is induced. Controls include Sex, Age, Education, Income, Religion, PV, Prosoc, and Numskill. Interactions indicates interaction terms of PV and Prosoc with ECOST. Robust standard errors are obtained by clustering at the individual level and appear in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

1.B.1.2 Payoff Implications

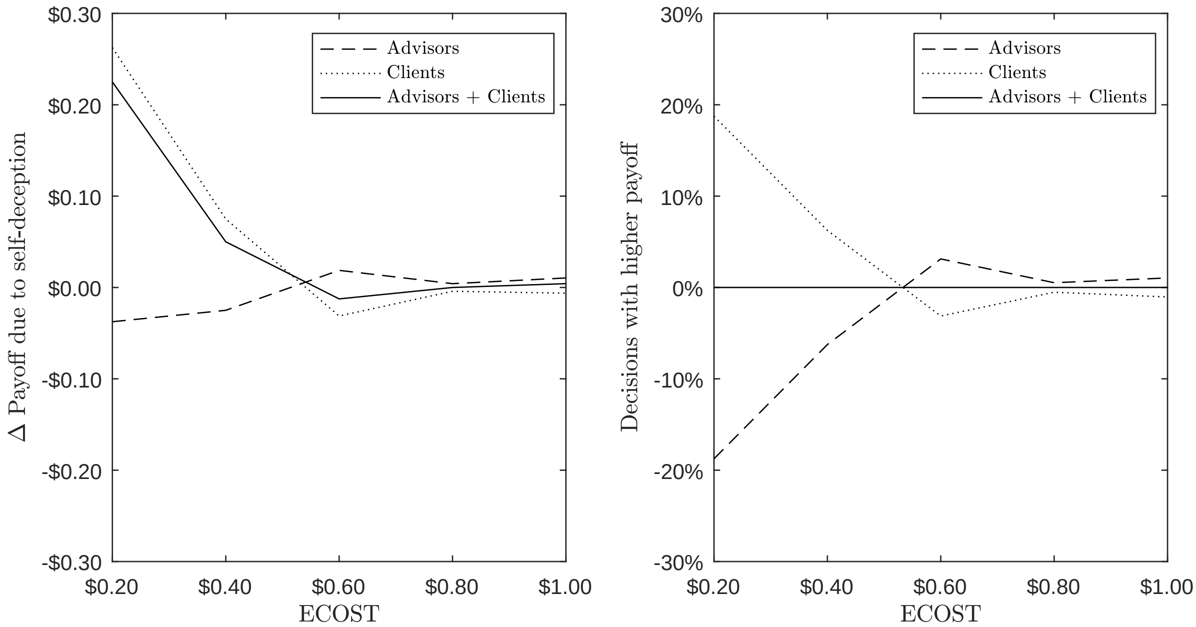


Figure 1.B.3: Payoff implications of self-deception in investment consulting based on stated beliefs. The left plot shows the differences in payoffs due to self-deception for advisors and clients. The right plot depicts the percentage of decisions in which a higher payoff is attained.

Table 1.B.3: Payoff implications in investment consulting based on stated beliefs

	Logit			OLS		
Dependent variable:	High Payoff AUT			Payoff AUT		
	(1)	(2)	(3)	(4)	(5)	(6)
ECOST	-1.027	-1.117	-2.330	39.820***	39.162***	24.241
	(1.099)	(1.087)	(2.143)	(13.460)	(13.335)	(22.880)
$ECOST^2$	0.706	0.767	0.793	-34.728***	-34.300***	-34.166***
	(0.872)	(0.862)	(0.871)	(11.394)	(11.278)	(11.302)
Fundamental bias	0.634	0.313	0.304	5.859	3.740	3.643
	(0.417)	(0.359)	(0.355)	(5.417)	(5.137)	(5.129)
$(\text{Fundamental bias})^2$		-0.415	-0.436		-2.845	-3.051
		(0.338)	(0.333)		(4.235)	(4.207)
Self-deceptive bias	0.330	0.327	0.343	5.926*	5.902*	6.063*
	(0.244)	(0.250)	(0.247)	(3.181)	(3.390)	(3.342)
$(\text{Self-deceptive bias})^2$		-0.049	-0.045		-0.388	-0.332
		(0.240)	(0.238)		(3.379)	(3.364)
Sex	0.217	0.207	0.208	2.396	2.301	2.308
	(0.315)	(0.313)	(0.314)	(4.136)	(4.119)	(4.122)
Age	-0.002	-0.003	-0.003	-0.038	-0.041	-0.041
	(0.019)	(0.019)	(0.019)	(0.242)	(0.242)	(0.242)
Education	-0.044	-0.055	-0.056	-0.797	-0.876	-0.880
	(0.232)	(0.232)	(0.232)	(2.960)	(2.944)	(2.942)
Income	0.092	0.101	0.102	0.905	0.971	0.971
	(0.121)	(0.121)	(0.122)	(1.568)	(1.573)	(1.574)
Religion	-0.270	-0.287	-0.287	-3.196	-3.313	-3.307
	(0.300)	(0.301)	(0.301)	(3.872)	(3.894)	(3.897)
PV	-0.664***	-0.677***	-0.732**	-8.928***	-9.001***	-10.053***
	(0.185)	(0.187)	(0.292)	(2.154)	(2.164)	(3.545)
ECOST × PV			0.089			1.751
			(0.285)			(2.982)
Prosoc	-0.081	-0.081	-0.227	-1.220	-1.214	-2.546
	(0.101)	(0.102)	(0.167)	(1.361)	(1.361)	(2.343)
ECOST × Prosoc			0.244			2.220
			(0.165)			(2.039)
Numskill	-0.049	-0.050	-0.050	0.161	0.163	0.171
	(0.140)	(0.139)	(0.139)	(1.743)	(1.741)	(1.742)
Constant	4.113***	4.263***	4.993***	160.679***	161.656***	170.559***
	(1.200)	(1.225)	(1.714)	(14.889)	(14.932)	(20.508)
Observations	960	960	960	960	960	960
Number of participants	96	96	96	96	96	96
R-squared	0.077	0.079	0.080	0.096	0.096	0.097

This table presents coefficients of logit regressions in columns (1), (2) and (3), and OLS regressions in columns (4), (5) and (6). The dependent variable in columns (1), (2) and (3) is High Payoff AUT which is a binary variable that takes the value one whenever the advisor ended up in the high payoff state in AUT, and zero else. The dependent variable in columns (4), (5) and (6) is the payoff in cents in each round from AUT. ECOST denotes economic costs of stating the truth, i.e., the monetary incentives to recommend a bad investment opportunity. The fundamental bias corresponds to EX:S−EX:A and the self-deceptive bias to AUT−EX:S. Sex, Education, Income, and Religion are as defined above. PV reflects an index of protected values for honesty. Prosoc captures prosocial concerns. Numskill is the number of correctly solved numerical patterns in the logical test. Robust standard errors are obtained by clustering at the individual level and appear in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

1.B.2 Neutral Setting

1.B.2.1 Bias Decomposition

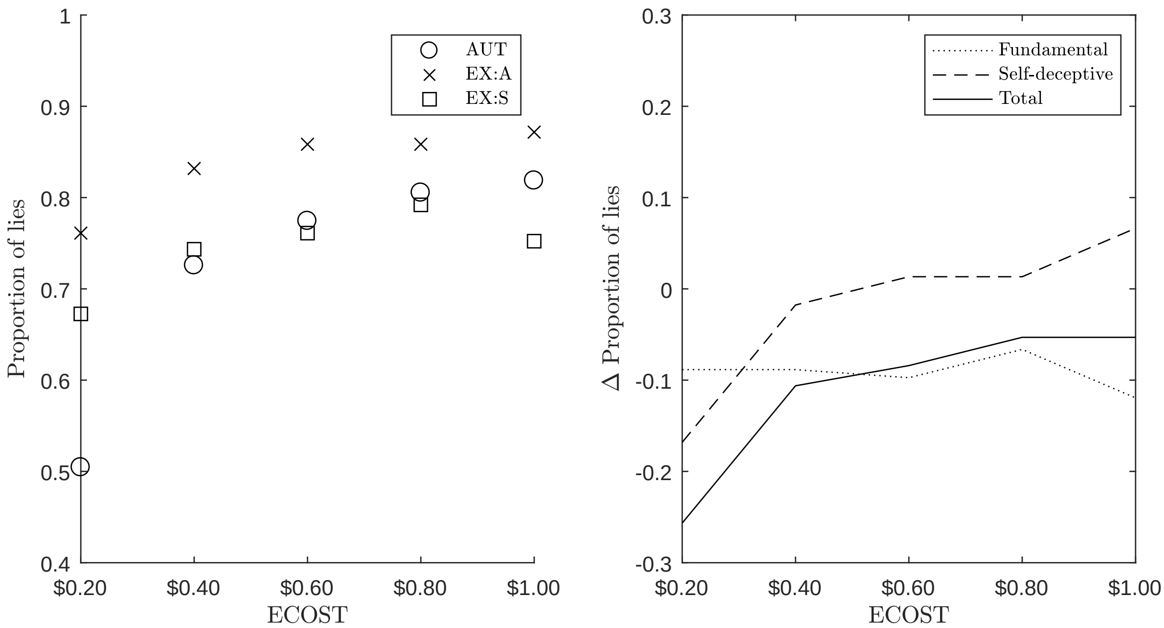


Figure 1.B.4: Lying in AUT, EX:A, and EX:S (left) and the bias decomposition in the neutral setting based on stated beliefs (right). Circles indicate the proportion of lies for given ECOST in AUT. Crosses correspond to the proportion of lies in EX:A, and squares denote the proportion of lies in EX:S. The fundamental bias is computed as EX:S−EX:A. The self-deceptive bias corresponds to AUT−EX:S. The total bias equals the sum of fundamental and self-deceptive bias.

Table 1.B.4: Determinants of the biases in the neutral setting based on stated beliefs

Dependent variable:	Total bias		Fundamental bias		Self-deceptive bias	
	(1)	(2)	(3)	(4)	(5)	(6)
ECOST	0.856***	1.060**	0.122	0.082	0.734***	0.977**
	(0.259)	(0.432)	(0.138)	(0.220)	(0.264)	(0.436)
$ECOST^2$	-0.521***	-0.521***	-0.119	-0.119	-0.403**	-0.403**
	(0.196)	(0.196)	(0.108)	(0.109)	(0.202)	(0.202)
Sex	0.030	0.030	0.083	0.083	-0.053	-0.053
	(0.055)	(0.055)	(0.055)	(0.055)	(0.067)	(0.067)
Age	-0.004	-0.004	-0.001	-0.001	-0.003	-0.003
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
Education	-0.023	-0.023	-0.022	-0.022	-0.001	-0.001
	(0.035)	(0.035)	(0.058)	(0.058)	(0.066)	(0.066)
Income	-0.003	-0.003	-0.004	-0.004	0.001	0.001
	(0.018)	(0.018)	(0.022)	(0.022)	(0.022)	(0.022)
Religion	0.010	0.010	0.005	0.005	0.004	0.004
	(0.058)	(0.058)	(0.058)	(0.058)	(0.066)	(0.066)
PV	0.003	0.027	-0.003	0.001	0.006	0.027
	(0.024)	(0.048)	(0.032)	(0.038)	(0.036)	(0.049)
ECOST × PV		-0.041		-0.007		-0.034
		(0.059)		(0.033)		(0.058)
Prosoc	-0.006	0.002	0.021	0.009	-0.027	-0.007
	(0.019)	(0.028)	(0.017)	(0.021)	(0.019)	(0.028)
ECOST × Prosoc		-0.014		0.019		-0.033
		(0.034)		(0.019)		(0.034)
Numskill	0.006	0.006	0.042**	0.042**	-0.036*	-0.036*
	(0.015)	(0.015)	(0.021)	(0.021)	(0.022)	(0.022)
Constant	-0.198	-0.320	-0.187	-0.163	-0.011	-0.157
	(0.243)	(0.337)	(0.272)	(0.294)	(0.291)	(0.350)
Observations	1,130	1,130	1,130	1,130	1,130	1,130
Number of participants	113	113	113	113	113	113
R-squared	0.039	0.039	0.041	0.041	0.040	0.041

This table presents coefficients of OLS regressions. The dependent variables are the total bias in columns (1) and (2), the fundamental bias in columns (3) and (4), and the self-deceptive bias in columns (5) and (6). The total bias is computed as AUT−EX:A, the fundamental bias as EX:S−EX:A, and the self-deceptive bias as AUT−EX:S. ECOST denotes economic costs of stating the truth, i.e., the monetary incentives to lie. Sex is a binary variable taking the value one if the participant reports to be male, and zero if female. Education takes the following values: 0: No schooling, 1: Elementary school, 2: High school, 3: Undergraduate degree, 4: Graduate school. Income reflects yearly household income and takes the following values: 1: $< \$20'000$, 2: $\$20'000 - 39'999$, 3: $\$40'000 - 59'999$, 4: $\$60'000 - 79'999$, 5: $\$80'000 - 99'999$, 6: $> \$100'000$. Religion denotes a binary variable which takes the value one if the participant indicates to belong either to Christianity, Islam, Judaism, Hinduism, Buddhism or a Folk religion, and zero else. Numskill is the number of correctly solved numerical patterns in the logical test. PV reflects an index of protected values for honesty. Prosoc captures prosocial concerns. Robust standard errors are obtained by clustering at the individual level and appear in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

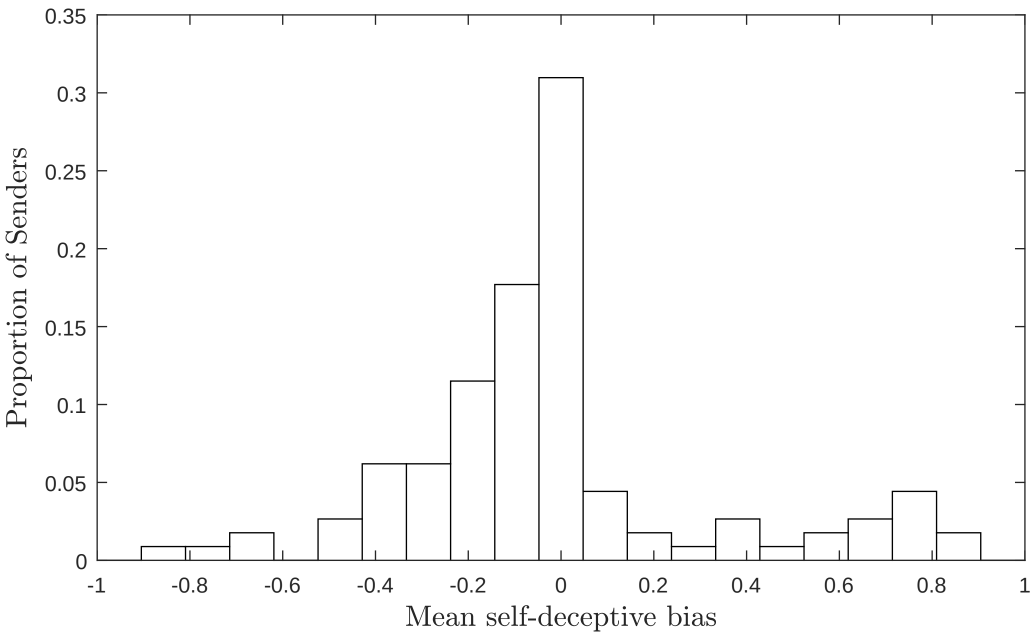


Figure 1.B.5: Cross-section of self-deceptive biases in the neutral setting based on stated beliefs. The self-deceptive bias is computed as the difference between behavior in AUT and EX:S averaged over all ECOST levels.

Table 1.B.5: Limiting self-deception in the neutral setting based on stated beliefs

Dependent variable:		Lie	
	(1)	(2)	(3)
ECOST	4.761***	5.372***	5.972***
	(0.701)	(0.739)	(1.514)
ECOST^2	-2.947***	-2.891***	-2.731***
	(0.526)	(0.528)	(0.531)
EX:S	0.107	0.864***	0.858***
	(0.190)	(0.241)	(0.235)
ECOST × EX:S		-1.342***	-1.349***
		(0.316)	(0.322)
Controls	Yes	Yes	Yes
Interactions	No	No	Yes
Observations	2,260	2,260	2,260
Number of participants	113	113	113
(Pseudo) R^2	0.091	0.097	0.102

This table presents coefficients of logit regressions. The dependent variable is the binary variable Lie which takes the value one if the Sender tells a lie, and zero else. ECOST denotes economic costs of stating the truth, i.e., the monetary incentives to lie. EX:S is a binary variable variable which takes the value one in EX:S where every Sender's stated belief is induced. Controls include Sex, Age, Education, Income, Religion, PV, Prosoc, and Numskill. Interactions indicates interaction terms of PV and Prosoc with ECOST. Robust standard errors are obtained by clustering at the individual level and appear in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

1.B.2.2 Payoff Implications

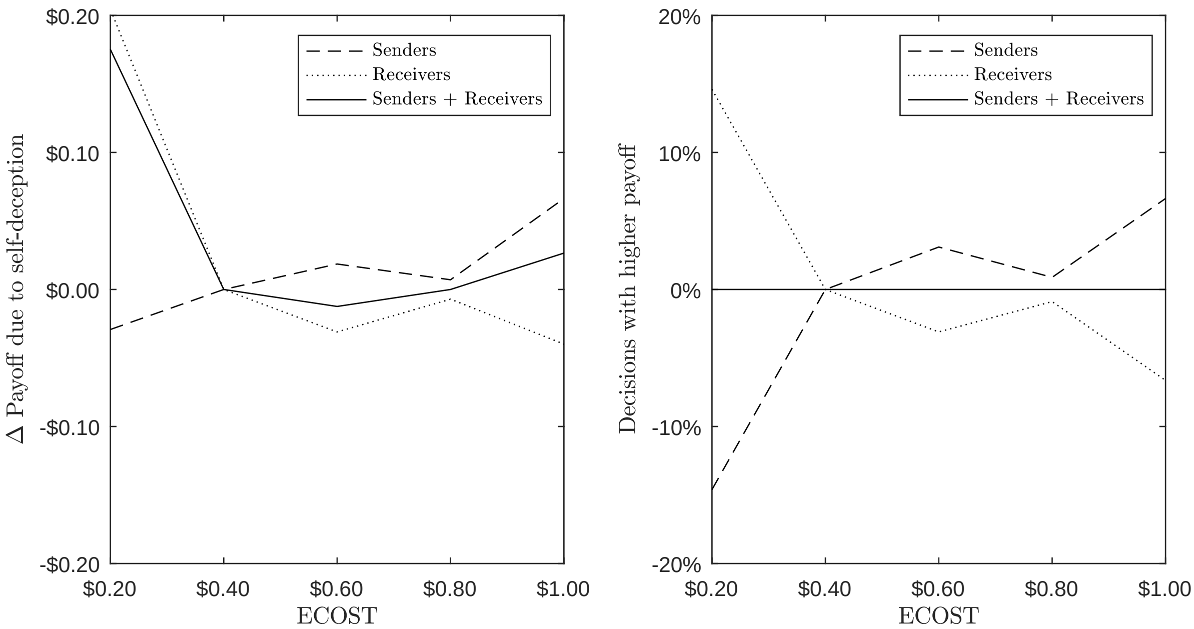


Figure 1.B.6: Payoff implications of self-deception in the neutral setting based on stated beliefs. The left plot shows the differences in payoffs due to self-deception for Senders and Receivers. The right plot depicts the percentage of decisions in which a higher payoff is attained.

Table 1.B.6: Payoff implications in the neutral setting based on stated beliefs

	Logit			OLS		
Dependent variable:	High Payoff AUT			Payoff AUT		
	(1)	(2)	(3)	(4)	(5)	(6)
ECOST	-3.234***	-3.322***	-3.745**	36.325***	35.554***	33.882**
	(1.234)	(1.259)	(1.846)	(11.938)	(12.105)	(16.467)
ECOST^2	2.168**	2.213**	2.223**	-34.641***	-34.285***	-34.315***
	(0.979)	(0.994)	(0.995)	(9.873)	(9.970)	(9.978)
Fundamental bias	0.387	0.184	0.188	6.085*	3.819	3.883
	(0.332)	(0.330)	(0.331)	(3.469)	(3.640)	(3.647)
(Fundamental bias)^2		-0.406	-0.407		-4.073	-4.103
		(0.334)	(0.335)		(3.528)	(3.540)
Self-deceptive bias	0.764***	0.781***	0.784***	8.717***	8.956***	8.978***
	(0.239)	(0.289)	(0.289)	(2.911)	(3.077)	(3.074)
(Self-deceptive bias)^2		-0.115	-0.111		-1.478	-1.380
		(0.297)	(0.299)		(3.112)	(3.142)
Sex	-0.200	-0.194	-0.194	-1.822	-1.777	-1.782
	(0.279)	(0.278)	(0.278)	(3.182)	(3.145)	(3.148)
Age	-0.033**	-0.032**	-0.032**	-0.334*	-0.325*	-0.325*
	(0.014)	(0.014)	(0.014)	(0.176)	(0.172)	(0.172)
Education	-0.077	-0.049	-0.049	-1.097	-0.852	-0.851
	(0.247)	(0.253)	(0.253)	(2.697)	(2.722)	(2.725)
Income	-0.038	-0.033	-0.033	-0.523	-0.465	-0.465
	(0.094)	(0.094)	(0.094)	(1.080)	(1.073)	(1.074)
Religion	-0.245	-0.217	-0.216	-2.457	-2.191	-2.192
	(0.281)	(0.277)	(0.277)	(3.173)	(3.116)	(3.120)
PV	-0.555***	-0.567***	-0.644**	-6.525**	-6.602**	-7.342**
	(0.208)	(0.211)	(0.295)	(2.558)	(2.571)	(3.460)
ECOST × PV			0.125			1.243
			(0.237)			(2.307)
Prosoc	-0.033	-0.037	-0.020	-0.096	-0.117	0.429
	(0.092)	(0.090)	(0.136)	(0.993)	(0.965)	(1.332)
ECOST × Prosoc			-0.028			-0.912
			(0.151)			(1.272)
Numskill	0.118	0.100	0.100	1.332	1.130	1.130
	(0.099)	(0.103)	(0.103)	(0.967)	(1.007)	(1.008)
Constant	5.924***	5.996***	6.250***	167.376***	167.917***	168.913***
	(1.182)	(1.210)	(1.531)	(13.502)	(13.855)	(16.663)
Observations	1,130	1,130	1,130	1,130	1,130	1,130
Number of participants	113	113	113	113	113	113
R-squared	0.076	0.079	0.080	0.078	0.081	0.081

This table presents coefficients of logit regressions in columns (1), (2) and (3), and OLS regressions in columns (4), (5) and (6). The dependent variable in columns (1), (2) and (3) is High Payoff AUT which is a binary variable that takes the value one whenever the Sender ended up in the high payoff state in AUT, and zero else. The dependent variable in columns (4), (5) and (6) is the payoff in cents in each round from AUT. ECOST denotes economic costs of stating the truth, i.e., the monetary incentives to lie. The fundamental bias corresponds to EX:S−EX:A and the self-deceptive bias to AUT−EX:S. Sex, Education, Income, and Religion are as defined above. PV reflects an index of protected values for honesty. Prosoc captures prosocial concerns. Numskill is the number of correctly solved numerical patterns in the logical test. Robust standard errors are obtained by clustering at the individual level and appear in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

1.C Appendix C: Additional Figures

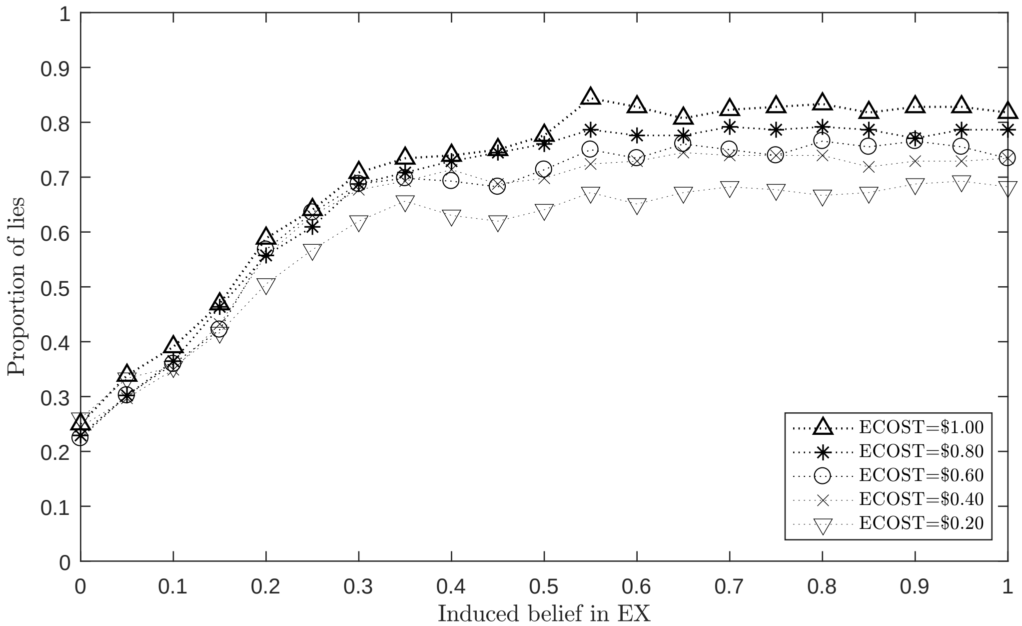


Figure 1.C.1: Unethical behavior across induced first-order beliefs in investment consulting. Upward-pointing triangles denote ECOST equal to $1.00, asterisks $0.80, circles $0.60, crosses $0.40, and downward-pointing triangles $0.20

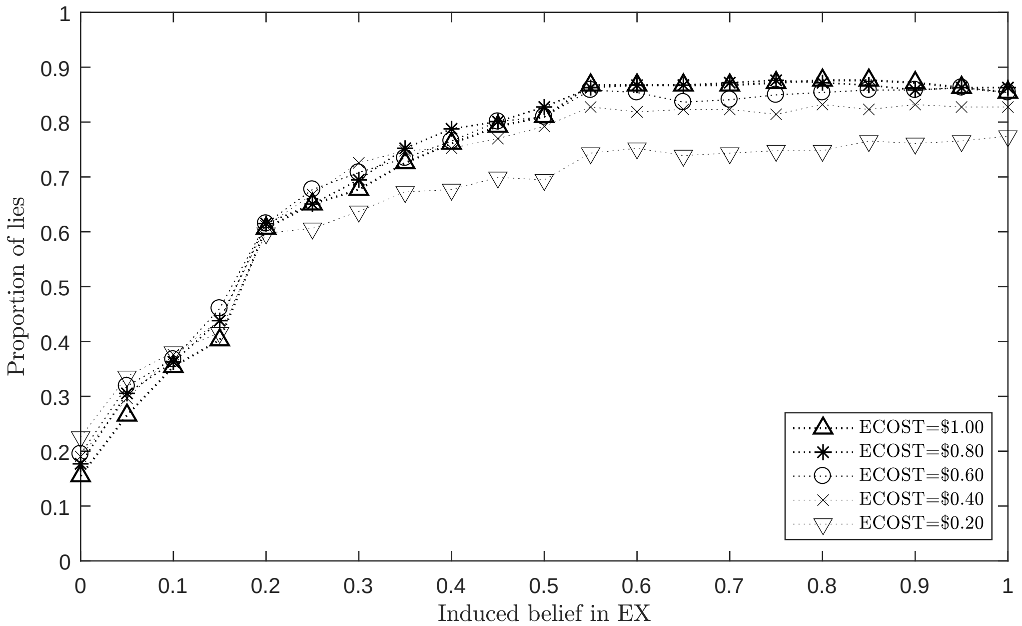


Figure 1.C.2: Unethical behavior across induced first-order beliefs in the neutral setting. Upward-pointing triangles denote ECOST equal to $1.00, asterisks $0.80, circles $0.60, crosses $0.40, and downward-pointing triangles $0.20

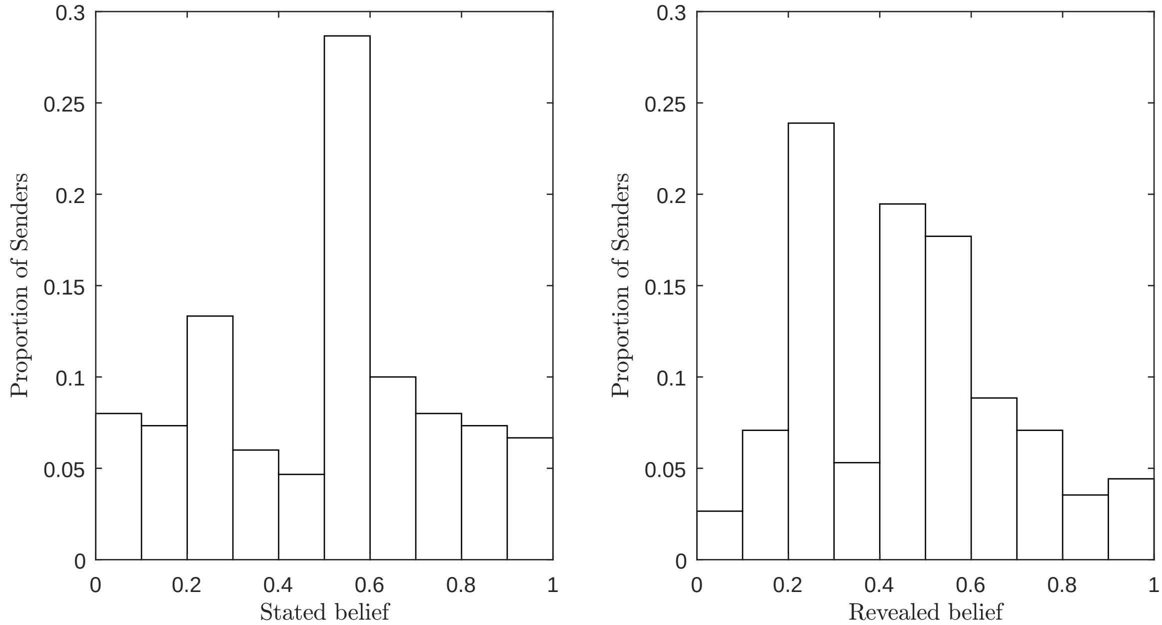


Figure 1.C.3: Histograms of stated (left) and revealed advisor beliefs in the neutral setting (right). Stated beliefs are unincentivized and self-reported. Revealed beliefs are derived according to the methodology of Andreoni and Sanchez (2014).

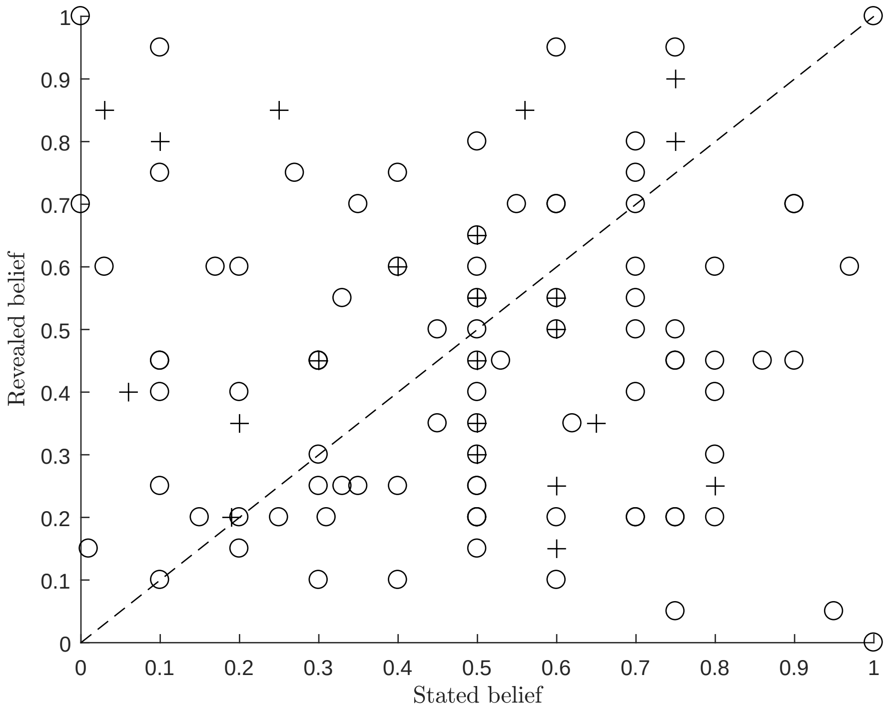


Figure 1.C.4: Stated versus revealed advisor beliefs in the neutral setting. Plus markers indicate participants who lie in at most 50% of all decisions. Circles mark participants who lie in more than 50% of all decisions.

1.D Appendix D: Experimental Instructions

1.D.1 Investment Consulting Experiment

1.D.1.1 Advisor

Introduction

In this study you will play a decision making experiment that includes several tasks. Please read the instructions carefully. You may earn a considerable sum of money, depending on the decisions you make in the experiment.

In this experiment you will earn experimental points which will be converted to real money at the end. The conversion rate is set such that 100 points in the experiment correspond to $1.00 (i.e., $0.01 per point in the experiment). The accumulated money you earn in these tasks will be paid to you at the end of the experiment in addition to a fixed amount of $0.25 for participation.

Keep in mind that you will only be paid if you fully complete the study until the end.

Task 1

In this task you will play 10 rounds of an investment advisory game. Please read the instructions carefully. You may earn a considerable sum of money, depending on the decisions you make.

In the end, one of the 10 rounds will be randomly selected and you will be randomly matched with another (anonymous) participant. Neither of you will know the identity of the other. The money you earn in this particular round will be paid to you at the end of the experiment in addition to the money you earn in the other tasks as well as a fixed amount of $0.25 for participation. The conversion rate is set such that 100 points in the experiment correspond to $1.00 (i.e., $0.01 per point in the experiment).

Instructions

You are in the role of an investment advisor who interacts with a client who is going to choose an investment. There are six possible investment opportunities. The other participant will not be informed about the characteristics of these investments. However, he or she will be told that you have been informed about the characteristics of the investments. In every round, you will be asked to send a message to the other participant. The message will correspond to one of the six investments. There are six possible messages:

- Message 1: "Investment 1 will yield the highest expected outcome."
- Message 2: "Investment 2 will yield the highest expected outcome."
- Message 3: "Investment 3 will yield the highest expected outcome."
- Message 4: "Investment 4 will yield the highest expected outcome."
- Message 5: "Investment 5 will yield the highest expected outcome."
- Message 6: "Investment 6 will yield the highest expected outcome."

Your message will be shown to the other participant, and then he or she will be asked to choose one investment.

The choice of the investment by the other participant will determine the payments in this task. For example the investments might be characterized as follows:

- Investment 1 has an expected outcome of 25 points, all other investments will yield an expected outcome of 32 points. The risk of all investments is the same.
- Your compensation will be 35 points if the other participant chooses Investment 1, and 28 points if he or she chooses any other investment.

Note that the characteristics of the investments as well as your possible compensation will change from round to round.

Finally, only you will be informed of the expected outcome of each particular investment. The other participant will not be informed. However, he or she will be told that you have been informed of the characteristics of each investment opportunity.

For your convenience, these instructions will remain available to you on all subsequent screens of this task.

On the next pages, you will have to answer two questions to make sure you understand the instructions correctly. Keep in mind that you will only be paid if you fully complete the study until the end.

Understanding Question [1-2] of 2

In this understanding question the investment opportunities are characterized as follows:

- Investment [ ] has an expected outcome of [Client Low Payoff] points, all other investments will yield an expected outcome of [Client High Payoff] points. The risk of all investments is the same.

- Your compensation will be [Advisor High Payoff] points if the other participant chooses Investment [ ] and [Advisor Low Payoff] points if he or she chooses any other investment.

Suppose you send the message: "Investment 5 will yield the highest expected outcome." and he or she chooses "Investment 2". Which of the following will be the result of that round?

Round [1-10] of 10

We ask you to send a message to the other participant. This message is the only information he or she will receive. He or she will neither be informed about the monetary consequences of each investment opportunity nor your compensation. In this round the monetary consequences are as follows:

- Investment [ ] has an expected outcome of [Client High Payoff] points, all other investments will yield an expected outcome of [Client Low Payoff] points. The risk of all investments is the same.
- Your compensation will be [Advisor Low Payoff] points if the other participant chooses Investment [ ], and [Advisor High Payoff] points if he or she chooses any other investment.

The message I will send is:

- Message 1: "Investment 1 will yield the highest expected outcome."
- Message 2: "Investment 2 will yield the highest expected outcome."
- Message 3: "Investment 3 will yield the highest expected outcome."
- Message 4: "Investment 4 will yield the highest expected outcome."
- Message 5: "Investment 5 will yield the highest expected outcome."
- Message 6: "Investment 6 will yield the highest expected outcome."

Question regarding your decisions

Out of 100 possible counterparts, how many do you think will follow your messages (i.e., choose the investment you mention in your message)?

Question regarding your preferred way of payment

Task 1 summary

The round randomly drawn to be played with an anonymous counterpart is: Round [ ]

In this round the investment opportunities are characterized as follows:

- Investment [ ] has an expected outcome of [Client High Payoff] points, all other investments will yield an expected outcome of [Client Low Payoff] points. The risk of all investments is the same.
- Your compensation will be [Advisor Low Payoff] points if the other participant chooses Investment [ ], and [Advisor High Payoff] points if he or she chooses any other investment.

You have sent the following message to the other player: Message [ ]: "Investment [ ] will yield the highest expected outcome."

We are now asking how you would like to be paid for this task. You have the choice between the outcome of the real game selected above whereas the other player has not yet chosen an investment or a lottery with known probabilities and the same two possible outcomes.

Below you will find a list of 21 possible lotteries. Please indicate for each lottery whether you prefer to receive the outcome from the selected round above (which is either [Advisor Low Payoff] or [Advisor High Payoff] points depending on the choice of the other player) or the outcome from the lottery. The lotteries are designed such that Lottery 1 will always pay you the highest possible outcome of the game, and Lottery 21 will always pay you less (in expectation) than the outcome of the game regardless of the other participant's decisions. The expected payoffs of Lotteries 2-20 lie between Lottery 1 (the best lottery) and Lottery 21 (the worst lottery).

As a result most people begin by preferring the lottery and then switch to the outcome of the game. Thus one way to view this task is to determine the best row to stop checking the box for the lottery and start checking the box for the outcome of the game. In the end one lottery corresponding to one row below will be randomly selected and you will be paid according to your indicated preference. The other player will receive the payment according to the game above regardless of your decision. Preferring the lottery will affect your payment only.

Lottery 1: It pays [Advisor Low Payoff] points with 0% probability (0 in 100) and [Advisor High Payoff] points with 100% probability (100 in 100).

• I prefer Lottery 1. • I prefer to obtain the outcome of the game above.

Lottery 2: It pays [Advisor Low Payoff] points with 5% probability (5 in 100) and [Advisor High Payoff] points with 95% probability (95 in 100).[47]

• I prefer Lottery 2. • I prefer to obtain the outcome of the game above.

Lottery 3: It pays [Advisor Low Payoff] points with 10% probability (10 in 100) and [Advisor High Payoff] points with 90% probability (90 in 100).

• I prefer Lottery 3. • I prefer to obtain the outcome of the game above.

Lottery 4: It pays [Advisor Low Payoff] points with 15% probability (15 in 100) and [Advisor High Payoff] points with 85% probability (85 in 100).

• I prefer Lottery 4. • I prefer to obtain the outcome of the game above.

Lottery 5: It pays [Advisor Low Payoff] points with 20% probability (20 in 100) and [Advisor High Payoff] points with 80% probability (80 in 100).

• I prefer Lottery 5. • I prefer to obtain the outcome of the game above.

Lottery 6: It pays [Advisor Low Payoff] points with 25% probability (25 in 100) and [Advisor High Payoff] points with 75% probability (75 in 100).

• I prefer Lottery 6. • I prefer to obtain the outcome of the game above.

Lottery 7: It pays [Advisor Low Payoff] points with 30% probability (30 in 100) and [Advisor High Payoff] points with 70% probability (70 in 100).

• I prefer Lottery 7. • I prefer to obtain the outcome of the game above.

Lottery 8: It pays [Advisor Low Payoff] points with 35% probability (35 in 100) and [Advisor High Payoff] points with 65% probability (65 in 100).

• I prefer Lottery 8. • I prefer to obtain the outcome of the game above.

Lottery 9: It pays [Advisor Low Payoff] points with 40% probability (40 in 100) and [Advisor High Payoff] points with 60% probability (60 in 100).

• I prefer Lottery 9. • I prefer to obtain the outcome of the game above.

Lottery 10: It pays [Advisor Low Payoff] points with 45% probability (45 in 100) and [Advisor High Payoff] points with 55% probability (55 in 100).

• I prefer Lottery 10. • I prefer to obtain the outcome of the game above.

Lottery 11: It pays [Advisor Low Payoff] points with 50% probability (50 in 100) and [Advisor High Payoff] points with 50% probability (50 in 100).

• I prefer Lottery 11. • I prefer to obtain the outcome of the game above.

Lottery 12: It pays [Advisor Low Payoff] points with 55% probability (55 in 100) and [Advisor High Payoff] points with 45% probability (45 in 100).

• I prefer Lottery 12. • I prefer to obtain the outcome of the game above.

Lottery 13: It pays [Advisor Low Payoff] points with 60% probability (60 in 100) and [Advisor High Payoff] points with 40% probability (40 in 100).

• I prefer Lottery 13. • I prefer to obtain the outcome of the game above.

Lottery 14: It pays [Advisor Low Payoff] points with 65% probability (65 in 100) and [Advisor High Payoff] points with 35% probability (35 in 100).

• I prefer Lottery 14. • I prefer to obtain the outcome of the game above.

Lottery 15: It pays [Advisor Low Payoff] points with 70% probability (70 in 100) and [Advisor High Payoff] points with 30% probability (30 in 100).

• I prefer Lottery 15. • I prefer to obtain the outcome of the game above.

Lottery 16: It pays [Advisor Low Payoff] points with 75% probability (75 in 100) and [Advisor High Payoff] points with 25% probability (25 in 100).

• I prefer Lottery 16. • I prefer to obtain the outcome of the game above.

Lottery 17: It pays [Advisor Low Payoff] points with 80% probability (80 in 100) and [Advisor High Payoff] points with 20% probability (20 in 100).

• I prefer Lottery 17. • I prefer to obtain the outcome of the game above.

Lottery 18: It pays [Advisor Low Payoff] points with 85% probability (85 in 100) and [Advisor High Payoff] points with 15% probability (15 in 100).

• I prefer Lottery 18. • I prefer to obtain the outcome of the game above.

Lottery 19: It pays [Advisor Low Payoff] points with 90% probability (90 in 100) and [Advisor High Payoff] points with 10% probability (10 in 100).

• I prefer Lottery 19. • I prefer to obtain the outcome of the game above.

Lottery 20: It pays [Advisor Low Payoff] points with 95% probability (95 in 100) and [Advisor High Payoff] points with 5% probability (5 in 100).

• I prefer Lottery 20. • I prefer to obtain the outcome of the game above.

Lottery 21: It pays [Advisor Low Payoff] points with 100% probability (100 in 100) and [Advisor High Payoff] points with 0% probability (0 in 100).

• I prefer Lottery 21. • I prefer to obtain the outcome of the game above.

[47] The probabilities depicted here correspond to the case when a round is drawn in which the advisor's message was truthful. If a round with a deceitful message is randomly chosen the probabilities are adjusted such that the lotteries again reflect the full range of possible first-order beliefs the advisor might have in the game.

Task 2

In this task you will have to complete numerical patterns. You will see 10 patterns and your task will be to determine the number that comes next in each series.

For each correctly solved pattern you will earn 10 points towards your total payment at the end of the experiment. The conversion rate is set such that 100 points in the experiment correspond to $1.00 (i.e., $0.01 per point in the experiment).

You have 120 seconds to determine as many missing numbers as possible. Once time is over you will be automatically redirected to the next page. You may start the clock by clicking on the button below.

Task 2

Time left to complete this page: [mm:ss]

Pattern 1: Determine which number comes next: 13, 17, 19, 23, 29, ?

Pattern 2: Determine which number comes next: 1, 4, 9, 16, 25, ?

Pattern 3: Determine which number comes next: 3, 18, 9, 54, 27, ?

Pattern 4: Determine which number comes next: 3, 4, 8, 11, 44, ?

Pattern 5: Determine which number comes next: 2, 5, 10, 17, 28, ?

Pattern 6: Determine which number comes next: 1, 1, 2, 6, 24, ?

Pattern 7: Determine which number comes next: 62, 54, 57, 49, 52, ?

Pattern 8: Determine which number comes next: 4, 12, 3, 9, 0, ?

Pattern 9: Determine which number comes next: 2, 4, 8, 16, 32, ?

Pattern 10: Determine which number comes next: 43, 32, 64, 53, 106, ?

Task 3

In this task you will play 10 rounds of an investment advisory game. In each round you have to make several decisions. Please read the instructions carefully. You may earn a considerable sum of money, depending on the decisions you make.

In the end, one of the 10 rounds will be randomly selected and you will be randomly matched with another (anonymous) participant. You will never be matched with the same participant as in Task 1. Neither of you will know the identity of the other. The money you earn in this particular round will be paid to you at the end of the experiment in addition to your earnings from the other tasks and a fixed amount for participation. The conversion rate is set such that 100 points in the experiment correspond to $1.00 (i.e., $0.01 per point in the experiment).

Instructions

You are in the role of an investment advisor who interacts with a client who is going to choose an investment. There are six possible investment opportunities. The other participant will not be informed about the characteristics of these investments. However, he or she will be told that you have been informed about the characteristics of the investments. In every round, you will be asked to send a message to the other participant. The message will correspond to one of the six investments. There are six possible messages:

- Message 1: "Investment 1 will yield the highest expected outcome."
- Message 2: "Investment 2 will yield the highest expected outcome."
- Message 3: "Investment 3 will yield the highest expected outcome."
- Message 4: "Investment 4 will yield the highest expected outcome."
- Message 5: "Investment 5 will yield the highest expected outcome."
- Message 6: "Investment 6 will yield the highest expected outcome."

Your message will be shown to the other participant, and then he or she will be asked to choose one investment.

The choice of the investment by the other participant will determine the payments in this task. For example the investments might be characterized as follows:

- Investment [ ] has an expected outcome of [Client High Payoff] points, all other investments will yield an expected outcome of [Client Low Payoff] points. The risk of all investments is the same.
- Your compensation will be [Advisor Low Payoff] points if the other participant chooses Investment [ ], and [Advisor High Payoff] points if he or she chooses any other investment.

Note that the characteristics of the investments as well as your possible compensation will change from round to round.

Again, only you will be informed of the expected outcome of each particular investment. The other participant will not be informed. However, he or she will be told that you have been informed of the characteristics of each investment opportunity.

In addition, you will always see a list of different participants with which you might be matched. You will know exactly how often these participants will choose the investment that you state in your message. You will have to indicate the message that you would like to send to each possible counterpart.

For your convenience, these instructions will remain available to you on all subsequent screens of this task.

Round [1-10] of 10

We ask you to send a message to all possible counterparts. This message is the only information he or she will receive. He or she will neither be informed about the monetary consequences of each investment opportunity nor your compensation. In this round the monetary consequences are as follows:

- Investment [ ] has an expected outcome of [Client High Payoff] points, all other investments will yield an expected outcome of [Client Low Payoff] points. The risk of all investments is the same.
- Your compensation will be [Advisor Low Payoff] points if the other participant chooses Investment [ ], and [Advisor High Payoff] points if he or she chooses any other investment.

Below you will find a list of the 21 participants with whom you might be matched. In addition there is information how each of them will play.

Please indicate for each participant which message you would like to send him or her. You will find the six possible messages which correspond to the six investments in the instructions below. In the end one participant will be randomly selected and you will be paid according to your decisions.

Participant 1 will choose the investment that you send in your message with 0% probability. Which message would you send Participant 1?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 2 will choose the investment that you send in your message with 5% probability. Which message would you send Participant 2?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 3 will choose the investment that you send in your message with 10% probability. Which message would you send Participant 3?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 4 will choose the investment that you send in your message with 15% probability. Which message would you send Participant 4?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 5 will choose the investment that you send in your message with 20% probability. Which message would you send Participant 5?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 6 will choose the investment that you send in your message with 25% probability. Which message would you send Participant 6?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 7 will choose the investment that you send in your message with 30% probability. Which message would you send Participant 7?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 8 will choose the investment that you send in your message with 35% probability. Which message would you send Participant 8?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 9 will choose the investment that you send in your message with 40% probability. Which message would you send Participant 9?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 10 will choose the investment that you send in your message with 45% probability. Which message would you send Participant 10?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 11 will choose the investment that you send in your message with 50% probability. Which message would you send Participant 11?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 12 will choose the investment that you send in your message with 55% probability. Which message would you send Participant 12?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 13 will choose the investment that you send in your message with 60% probability. Which message would you send Participant 13?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 14 will choose the investment that you send in your message with 65% probability. Which message would you send Participant 14?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 15 will choose the investment that you send in your message with 70% probability. Which message would you send Participant 15?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 16 will choose the investment that you send in your message with 75% probability. Which message would you send Participant 16?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 17 will choose the investment that you send in your message with 80% probability. Which message would you send Participant 17?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 18 will choose the investment that you send in your message with 85% probability. Which message would you send Participant 18?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 19 will choose the investment that you send in your message with 90% probability. Which message would you send Participant 19?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 20 will choose the investment that you send in your message with 95% probability. Which message would you send Participant 20?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 21 will choose the investment that you send in your message with 100% probability. Which message would you send Participant 21?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Questionnaire

Question 1: Please indicate your gender:

• Female • Male

Question 2: Please indicate your age:

Question 3: What is the highest degree/level of education you have completed?

• No schooling • Elementary school • Middle school • High school • Undergraduate school • Graduate school

Question 4: What is your average yearly household income (in USD)?

• < 20‘000 • 20‘001 - 40‘000 • 40‘001 - 60‘000 • 60‘001 - 80‘000 • 80‘001 - 100‘000 • > 100‘000 • I do not want to say

Question 5: To which religion do you belong?

• Christianity • Islam • Hinduism • Buddhism • Judaism • Folk religion • No religion • Other

Question 6: Suppose that Investment 5 yields the highest expected outcome. To what extent do you judge the decision to send something other than Message 5: "Investment 5 will yield the highest expected outcome." as:[48]

[48] The measure Prosoc is computed based on subquestions 2 and 4 while 1 and 3 serve as consistency check.

- 1: Very dishonest ... 7: Very honest
- 1: Very manipulative ... 7: Not at all manipulative
- 1: Strongly associated with personal gains ... 7: Not at all associated with personal gains
- 1: Strongly hurting the other participant ... 7: Not at all hurting the other participant

Question 7: Consider the choices of the player who sent a message. To what extent do you judge the decision to send something other than the truthful message as:[49]

- 1: Very immoral ... 7: Very moral
- 1: Not at all praiseworthy ... 7: Very praiseworthy
- 1: Not at all blameworthy ... 7: Very blameworthy
- 1: Not at all outrageous ... 7: Very outrageous
- 1: Not at all acceptable ... 7: Very acceptable

Question 8: Individuals have an opportunity to choose which information they provide to their audience, and they can modify information they obtain before passing this information on to others. Some view such modification as a violation of truthfulness; others regard it as acceptable protection of personal interests. What do you think about the value of truthfulness in such a situation? Truthfulness is about something...

- ... that one should not sacrifice, no matter what the (material or other) benefits. [1: Strongly Disagree ... 7: Strongly Agree]
- ... for which I think it is right to make a cost-benefit analysis. [1: Strongly Disagree ... 7: Strongly Agree]

[49] In Questions 7 and 8, I restate some variables such that a large number always corresponds to strong protected values for honesty.

- ... that cannot be measured in monetary terms. [1: Strongly Disagree ... 7:
- ... about which I can be flexible if the situation demands it. [1: Strongly Disagree ... 7: Strongly Agree]

Thank you!

You have completed the experiment session. Thank you for doing our study!

Very soon you will receive the fixed fee of $0.25. In the next hours we will randomly match you with a player for Task 1 and another player for Task 3 and also pay the bonus that you have earned depending on the rounds selected by the lotteries as well as the other participants' choices. In addition you will receive the money you have earned in Task 2. We usually manage to pay participants within 24 hours.

Do you have any comments or feedback regarding this experiment?

• Yes • No

1.D.1.2 Client

Introduction

In this study you will play a decision making experiment that includes three tasks. Please read the instructions carefully. You may earn a considerable sum of money, depending on the decisions you make in the experiment.

In this experiment you will earn experimental points which will be converted to real money at the end. The conversion rate is set such that 100 points in the experiment correspond to $1.00 (i.e., $0.01 per point in the experiment). The money you earn in these tasks will be paid to you at the end of the experiment in addition to a fixed amount of $0.25 for participation.

Keep in mind that you will only be paid if you fully complete the study until the end.

Task 1

Instructions

In this task you will play a short decision making game. Please read the instructions carefully. You may earn a considerable sum of money, depending on the decisions you make in the experiment.

You will be randomly matched with another (anonymous) participant in this task. Neither of you will know the identity of the other.

There are six possible investment opportunities. The other participant knows the expected outcome of all of them, but we are not going to tell it to you.

After being informed about the characteristics of the six investments, the other participant will send a message to you. The message corresponds to a recommendation for one of the investments. There are six possible messages:

- Message 1: "Investment 1 will yield the highest expected outcome."
- Message 2: "Investment 2 will yield the highest expected outcome."
- Message 3: "Investment 3 will yield the highest expected outcome."
- Message 4: "Investment 4 will yield the highest expected outcome."
- Message 5: "Investment 5 will yield the highest expected outcome."
- Message 6: "Investment 6 will yield the highest expected outcome."

We will ask you to choose one of the six investments. The message you receive is the only information you will have regarding the expected outcomes. Your choice of an investment will determine the payments according to two different options, known only to the other participant.

One investment will pay you more than the other investments. For your convenience, these instructions will remain available to you on all subsequent screens of this task.

Decisions

The other participant will send you a message. For every possible message we ask you to choose one of the six investment opportunities. Your choice will determine the payments to both of you.

Decision 1: Suppose the other participant sent you: Message 1: "Investment 1 will yield the highest expected outcome."

Given this message the investment I choose is:

• Investment 1 • Investment 2 • Investment 3 • Investment 4 • Investment 5 • Investment 6

Decision 2: Suppose the other participant sent you: Message 2: "Investment 2 will yield the highest expected outcome."

Given this message the investment I choose is:

• Investment 1 • Investment 2 • Investment 3 • Investment 4 • Investment 5 • Investment 6

Decision 3: Suppose the other participant sent you: Message 3: "Investment 3 will yield the highest expected outcome."

Given this message the investment I choose is:

• Investment 1 • Investment 2 • Investment 3 • Investment 4 • Investment 5 • Investment 6

Decision 4: Suppose the other participant sent you: Message 4: "Investment 4 will yield the highest expected outcome."

Given this message the investment I choose is:

• Investment 1 • Investment 2 • Investment 3 • Investment 4 • Investment 5 • Investment 6

Decision 5: Suppose the other participant sent you: Message 5: "Investment 5 will yield the highest expected outcome."

Given this message the investment I choose is:

• Investment 1 • Investment 2 • Investment 3 • Investment 4 • Investment 5 • Investment 6

Decision 6: Suppose the other participant sent you: Message 6: "Investment 6 will yield the highest expected outcome."

Given this message the investment I choose is:

• Investment 1 • Investment 2 • Investment 3 • Investment 4 • Investment 5 • Investment 6

Question regarding your decisions

Out of 100 possible counterparts, how many do you think will send you a truthful message (i.e., send the message with the investment that yields the highest expected outcome)?

Task 2

In this task you will have to complete numerical patterns. You will see 10 patterns and your task will be to determine the number that comes next in each series.

For each correctly solved pattern you will earn 10 points towards your total payment at the end of the experiment. The conversion rate is set such that 100 points in the experiment correspond to $1.00 (i.e., $0.01 per point in the experiment).

You have 120 seconds to determine as many missing numbers as possible. Once time is over you will be automatically redirected to the next page. You may start the clock by clicking on the button below.

This task is followed by some general questions.

Task 2

Time left to complete this page: [mm:ss]

Pattern 1: Determine which number comes next: 13, 17, 19, 23, 29, ?

Pattern 2: Determine which number comes next: 1, 4, 9, 16, 25, ?

Pattern 3: Determine which number comes next: 3, 18, 9, 54, 27, ?

Pattern 4: Determine which number comes next: 3, 4, 8, 11, 44, ?

Pattern 5: Determine which number comes next: 2, 5, 10, 17, 28, ?

Pattern 6: Determine which number comes next: 1, 1, 2, 6, 24, ?

Pattern 7: Determine which number comes next: 62, 54, 57, 49, 52, ?

Pattern 8: Determine which number comes next: 4, 12, 3, 9, 0, ?

Pattern 9: Determine which number comes next: 2, 4, 8, 16, 32, ?

Pattern 10: Determine which number comes next: 43, 32, 64, 53, 106, ?

Task 3

Instructions

Please read the instructions carefully. You may earn a considerable sum of money, depending on the decisions you make in this task.

In this task you will face decision situations with 10 anonymous participants. Neither of you will know the identity of the other.

There are six possible investment opportunities. The other participant knows the expected outcome of all of them, but we are not going to tell it to you.

After being informed about the characteristics of the six investments, the other participant will send a message to you. The message corresponds to a recommendation for one of the investments. There are six possible messages:

- Message 1: "Investment 1 will yield the highest expected outcome."
- Message 2: "Investment 2 will yield the highest expected outcome."
- Message 3: "Investment 3 will yield the highest expected outcome."
- Message 4: "Investment 4 will yield the highest expected outcome."
- Message 5: "Investment 5 will yield the highest expected outcome."
- Message 6: "Investment 6 will yield the highest expected outcome."

The message you receive is the only information you will have regarding the expected outcomes. We will ask you to choose one of the six investments. In addition, you are forced to choose the investment that was stated in the message exactly [ ] times in total (out of 60 decisions).

In the end a lottery will randomly determine with which of the 10 participants you will be matched. Your choice of an investment will determine the payments according to two different options, known only to the other participant. One investment will pay you more than the other investments.

For your convenience, these instructions will remain available to you on all subsequent screens of this task.

Decisions for participant [1-10]

Participant [1-10] will send you a message. For every possible message we ask you to choose one of the six numbers. Your choice will determine the payments to both of you.

Keep in mind that you must choose the same number as the one in the message in exactly [ ] out of the 60 cases in total.

You have already chosen a number equal to the number stated in the message [ ] times. This means that you must do so in [ ] more cases.

Decision 1: Suppose the other participant sent you: Message 1: "Investment 1 will yield the highest expected outcome."

Given this message the number I choose is:

• Investment 1 • Investment 2 • Investment 3 • Investment 4 • Investment 5 • Investment 6

Decision 2: Suppose the other participant sent you: Message 2: "Investment 2 will yield the highest expected outcome."

Given this message the number I choose is:

• Investment 1 • Investment 2 • Investment 3 • Investment 4 • Investment 5 • Investment 6

Decision 3: Suppose the other participant sent you: Message 3: "Investment 3 will yield the highest expected outcome."

Given this message the number I choose is:

• Investment 1 • Investment 2 • Investment 3 • Investment 4 • Investment 5 • Investment 6

Decision 4: Suppose the other participant sent you: Message 4: "Investment 4 will yield the highest expected outcome."

Given this message the number I choose is:

• Investment 1 • Investment 2 • Investment 3 • Investment 4 • Investment 5 • Investment 6

Decision 5: Suppose the other participant sent you: Message 5: "Investment 5 will yield the highest expected outcome."

Given this message the number I choose is:

• Investment 1 • Investment 2 • Investment 3 • Investment 4 • Investment 5 • Investment 6

Decision 6: Suppose the other participant sent you: Message 6: "Investment 6 will yield the highest expected outcome."

Given this message the number I choose is:

• Investment 1 • Investment 2 • Investment 3 • Investment 4 • Investment 5 • Investment 6

Questionnaire

Question 1: Please indicate your gender:

• Female • Male

Question 2: Please indicate your age:

Question 3: What is the highest degree/level of education you have completed?

• No schooling • Elementary school • Middle school • High school • Undergraduate school • Graduate school

Question 4: What is your average yearly household income (in USD)?

• < 20‘000 • 20‘001 - 40‘000 • 40‘001 - 60‘000 • 60‘001 - 80‘000 • 80‘001 - 100‘000 • > 100‘000 • I do not want to say

Question 5: To which religion do you belong?

• Christianity • Islam • Hinduism • Buddhism • Judaism • Folk religion • No religion • Other

Question 6: Suppose Investment 5 yields the highest expected outcome. To what extent do you judge the decision to send something other than Message 5: "Investment 5 will yield the highest expected outcome." as:

- 1: Very dishonest ... 7: Very honest
- 1: Very manipulative ... 7: Not at all manipulative
- 1: Strongly associated with personal gains ... 7: Not at all associated with personal gains
- 1: Strongly hurting the other participant ... 7: Not at all hurting the other participant

Question 7: Consider the choices of the player who sent you a message. To what extent do you judge the decision to send something other than the truthful message as:

- 1: Very immoral ... 7: Very moral
- 1: Not at all praiseworthy ... 7: Very praiseworthy
- 1: Not at all blameworthy ... 7: Very blameworthy
- 1: Not at all outrageous ... 7: Very outrageous
- 1: Not at all acceptable ... 7: Very acceptable

Question 8: Individuals have an opportunity to choose which information they provide to their audience, and they can modify information they obtain before passing this information on to others. Some view such modification as a violation of truthfulness; others regard it as acceptable protection of personal interests. What do you think about the value of truthfulness in such a situation? Truthfulness is about something...

- ... that one should not sacrifice, no matter what the (material or other) benefits. [1: Strongly Disagree ... 7: Strongly Agree]
- ... for which I think it is right to make a cost-benefit analysis. [1: Strongly Disagree ... 7: Strongly Agree]

- ... that cannot be measured in monetary terms. [1: Strongly Disagree ... 7:
- ... about which I can be flexible if the situation demands it. [1: Strongly Disagree ... 7: Strongly Agree]

Thank you!

You have completed the experiment. Thank you for doing our study!

Very soon you will receive the fixed fee of $0.25. In the next hours we will randomly match you with a player for Task 1 and another player for Task 3 and also pay the bonus that you have earned depending on the rounds selected by the lotteries as well as the other player's choice. In addition you will receive the money you have earned in Task 2. We usually manage to pay participants within 24 hours.

Do you have any comments or feedback regarding this experiment?

• Yes • No

1.D.2 Neutral Setting

1.D.2.1 Sender

Introduction

In this study you will play a decision making experiment that includes several tasks. Please read the instructions carefully. You may earn a considerable sum of money, depending on the decisions you make in the experiment.

In this experiment you will earn experimental points which will be converted to real money at the end. The conversion rate is set such that 100 points in the experiment correspond to $1.00 (i.e., $0.01 per point in the experiment). The accumulated money you earn in these tasks will be paid to you at the end of the experiment in addition to a fixed amount of $0.25 for participation.

Keep in mind that you will only be paid if you fully complete the study until the end.

Task 1

In this task you will play 10 rounds of a short decision making game. Please read the instructions carefully. You may earn a considerable sum of money, depending on the decisions you make.

In the end, one of the 10 rounds will be randomly selected and you will be randomly matched with another (anonymous) participant. Neither of you will know the identity of the other. The money you earn in this particular round will be paid to you at the end of the experiment in addition to the money you earn in the other tasks as well as a fixed amount of $0.25 for participation. The conversion rate is set such that 100 points in the experiment correspond to $1.00 (i.e., $0.01 per point in the experiment).

Instructions

Before every round, we will roll a 6-sided die. The other participant will not be informed about the outcome of the die roll. However, he or she will be told that you have been informed about the outcome of the die roll. In every round, you will be asked to send a message to the other participant. The message will correspond to a number from 1 to 6. There are six possible messages:

- Message 1: "The outcome from the roll of the 6-sided die is 1."
- Message 2: "The outcome from the roll of the 6-sided die is 2."
- Message 3: "The outcome from the roll of the 6-sided die is 3."

- Message 4: "The outcome from the roll of the 6-sided die is 4."
- Message 5: "The outcome from the roll of the 6-sided die is 5."
- Message 6: "The outcome from the roll of the 6-sided die is 6."

Your message will be shown to the other participant, and then he or she will be asked to choose a number between 1 and 6.

The choice of the number by the other participant will determine the payments in this task. For example:

- If he or she chooses the actual outcome of the roll of the die, then you will receive 35 points and he or she will receive 25 points.
- If he or she chooses a number different than the actual outcome, you will receive 28 points and he or she will receive 32 points.

Note that the monetary payments will change from round to round.

Finally, only you will be informed of the particular monetary value connected to each number. The other participant will not be informed of these monetary values. However, he or she will be told that you have been informed of the monetary value connected to each number.

For your convenience, these instructions will remain available to you on all subsequent screens of this task.

On the next pages, you will have to answer two questions to make sure you understand the instructions correctly. Keep in mind that you will only be paid if you fully complete the study until the end.

Understanding Question [1-2] of 2

In this understanding question the monetary consequences are as follows:

- If he or she chooses the actual outcome of the roll of the die, then you will receive [Sender Low Payoff] points and he or she will receive [Receiver High Payoff] points.
- If he or she chooses a number different than the actual outcome of the roll of the die, you will receive [Sender High Payoff] points and he or she will receive [Receiver Low Payoff] points.

The actual outcome of the die roll, which was conducted before starting this example round, is "2".

Suppose you send the message: "The outcome from the roll of the 6-sided die is 5." and he or she chooses "2".

Which of the following will be the result of that round?

Round [1-10] of 10

Before starting this round, we have rolled a 6-sided die, and obtained the outcome 3. We ask you to send a message to the other participant. This message is the only information he or she will receive. He or she will not be informed about the monetary consequences of each option. In this round the monetary consequences are as follows:

- If he or she chooses the actual outcome of the roll of the die, then you will receive [Sender Low Payoff] points and he or she will receive [Receiver High Payoff] points.
- If he or she chooses a number different than the actual outcome of the roll of the die, you will receive [Sender High Payoff] points and he or she will receive [Receiver Low Payoff] points.

The message I will send is:

- Message 1: "The outcome from the roll of the 6-sided die is 1."
- Message 2: "The outcome from the roll of the 6-sided die is 2."
- Message 3: "The outcome from the roll of the 6-sided die is 3."
- Message 4: "The outcome from the roll of the 6-sided die is 4."
- Message 5: "The outcome from the roll of the 6-sided die is 5."
- Message 6: "The outcome from the roll of the 6-sided die is 6."

Question regarding your decisions

Out of 100 possible counterparts, how many do you think will follow your messages (i.e., choose the number you mention in your message)?

Question regarding your preferred way of payment

Task 1 summary

The round randomly drawn to be played with an anonymous counterpart is: Round [ ]

In this round the actual outcome of the 6-sided die is: [ ] The associated monetary consequences are:

- If he or she chooses the actual outcome of the roll of the die, then you will receive [Sender Low Payoff] points and he or she will receive [Receiver High Payoff] points.
- If he or she chooses a number different than the actual outcome of the roll of the die, you will receive [Sender High Payoff] points and he or she will receive [Receiver Low Payoff] points.

You have sent the following message to the other player: Message [ ]: "The outcome from the roll of the 6-sided die is [ ]."

We are now asking how you would like to be paid for this task. You have the choice between the outcome of the real game selected above whereas the other player has not yet chosen a number or a lottery with known probabilities and the same two possible outcomes.

Below you will find a list of 21 possible lotteries. Please indicate for each lottery whether you prefer to receive the outcome from the selected round above (which is either [Sender Low Payoff] or [Sender High Payoff] points depending on the action of the other player) or the outcome from the lottery. The lotteries are designed such that Lottery 1 will always pay you the highest possible outcome of the game, and Lottery 21 will always pay you less (in expectation) than the outcome of the game regardless of the other participant's decisions. The expected payoffs of Lotteries 2-20 lie between Lottery 1 (the best lottery) and Lottery 21 (the worst lottery).

As a result most people begin by preferring the lottery and then switch to the outcome of the game. Thus one way to view this task is to determine the best row to stop checking the box for the lottery and start checking the box for the outcome of the game. In the end one lottery corresponding to one row below will be randomly selected and you will be paid according to your indicated preference. The other player will receive the payment according to the game above regardless of your decision. Preferring the lottery will affect your payment only.

Lottery 1: It pays [Sender Low Payoff] points with 0% probability (0 in 100) and [Sender High Payoff] points with 100% probability (100 in 100).

• I prefer Lottery 1. • I prefer to obtain the outcome of the game above.

Lottery 2: It pays [Sender Low Payoff] points with 5% probability (5 in 100) and [Sender High Payoff] points with 95% probability (95 in 100).[50]

• I prefer Lottery 2. • I prefer to obtain the outcome of the game above.

Lottery 3: It pays [Sender Low Payoff] points with 10% probability (10 in 100) and [Sender High Payoff] points with 90% probability (90 in 100).

• I prefer Lottery 3. • I prefer to obtain the outcome of the game above.

Lottery 4: It pays [Sender Low Payoff] points with 15% probability (15 in 100) and [Sender High Payoff] points with 85% probability (85 in 100).

• I prefer Lottery 4. • I prefer to obtain the outcome of the game above.

Lottery 5: It pays [Sender Low Payoff] points with 20% probability (20 in 100) and [Sender High Payoff] points with 80% probability (80 in 100).

• I prefer Lottery 5. • I prefer to obtain the outcome of the game above.

Lottery 6: It pays [Sender Low Payoff] points with 25% probability (25 in 100) and [Sender High Payoff] points with 75% probability (75 in 100).

• I prefer Lottery 6. • I prefer to obtain the outcome of the game above.

Lottery 7: It pays [Sender Low Payoff] points with 30% probability (30 in 100) and [Sender High Payoff] points with 70% probability (70 in 100).

• I prefer Lottery 7. • I prefer to obtain the outcome of the game above.

Lottery 8: It pays [Sender Low Payoff] points with 35% probability (35 in 100) and [Sender High Payoff] points with 65% probability (65 in 100).

• I prefer Lottery 8. • I prefer to obtain the outcome of the game above.

Lottery 9: It pays [Sender Low Payoff] points with 40% probability (40 in 100) and [Sender High Payoff] points with 60% probability (60 in 100).

• I prefer Lottery 9. • I prefer to obtain the outcome of the game above.

Lottery 10: It pays [Sender Low Payoff] points with 45% probability (45 in 100) and [Sender High Payoff] points with 55% probability (55 in 100).

• I prefer Lottery 10. • I prefer to obtain the outcome of the game above.

Lottery 11: It pays [Sender Low Payoff] points with 50% probability (50 in 100) and [Sender High Payoff] points with 50% probability (50 in 100).

• I prefer Lottery 11. • I prefer to obtain the outcome of the game above.

[50] The probabilities depicted here correspond to the case when a round is drawn in which the Sender's message was truthful. If a round with a deceitful message is randomly chosen the probabilities are adjusted such that the lotteries again reflect the full range of possible first-order beliefs the Sender might have in the game.

Lottery 12: It pays [Sender Low Payoff] points with 55% probability (55 in 100) and [Sender High Payoff] points with 45% probability (45 in 100).

• I prefer Lottery 12. • I prefer to obtain the outcome of the game above.

Lottery 13: It pays [Sender Low Payoff] points with 60% probability (60 in 100) and [Sender High Payoff] points with 40% probability (40 in 100).

• I prefer Lottery 13. • I prefer to obtain the outcome of the game above.

Lottery 14: It pays [Sender Low Payoff] points with 65% probability (65 in 100) and [Sender High Payoff] points with 35% probability (35 in 100).

• I prefer Lottery 14. • I prefer to obtain the outcome of the game above.

Lottery 15: It pays [Sender Low Payoff] points with 70% probability (70 in 100) and [Sender High Payoff] points with 30% probability (30 in 100).

• I prefer Lottery 15. • I prefer to obtain the outcome of the game above.

Lottery 16: It pays [Sender Low Payoff] points with 75% probability (75 in 100) and [Sender High Payoff] points with 25% probability (25 in 100).

• I prefer Lottery 16. • I prefer to obtain the outcome of the game above.

Lottery 17: It pays [Sender Low Payoff] points with 80% probability (80 in 100) and [Sender High Payoff] points with 20% probability (20 in 100).

• I prefer Lottery 17. • I prefer to obtain the outcome of the game above.

Lottery 18: It pays [Sender Low Payoff] points with 85% probability (85 in 100) and [Sender High Payoff] points with 15% probability (15 in 100).

• I prefer Lottery 18. • I prefer to obtain the outcome of the game above.

Lottery 19: It pays [Sender Low Payoff] points with 90% probability (90 in 100) and [Sender High Payoff] points with 10% probability (10 in 100).

• I prefer Lottery 19. • I prefer to obtain the outcome of the game above.

Lottery 20: It pays [Sender Low Payoff] points with 95% probability (95 in 100) and [Sender High Payoff] points with 5% probability (5 in 100).

• I prefer Lottery 20. • I prefer to obtain the outcome of the game above.

Lottery 21: It pays [Sender Low Payoff] points with 100% probability (100 in 100) and [Sender High Payoff] points with 0% probability (0 in 100).

• I prefer Lottery 21. • I prefer to obtain the outcome of the game above.

Task 2

In this task you will have to complete numerical patterns. You will see 10 patterns and your task will be to determine the number that comes next in each series.

For each correctly solved pattern you will earn 10 points towards your total payment at the end of the experiment. The conversion rate is set such that 100 points in the experiment correspond to $1.00 (i.e., $0.01 per point in the experiment).

You have 120 seconds to determine as many missing numbers as possible. Once time is over you will be automatically redirected to the next page. You may start the clock by clicking on the button below.

Task 2

Time left to complete this page: [mm:ss]

Pattern 1: Determine which number comes next: 13, 17, 19, 23, 29, ?

Pattern 2: Determine which number comes next: 1, 4, 9, 16, 25, ?

Pattern 3: Determine which number comes next: 3, 18, 9, 54, 27, ?

Pattern 4: Determine which number comes next: 3, 4, 8, 11, 44, ?

Pattern 5: Determine which number comes next: 2, 5, 10, 17, 28, ?

Pattern 6: Determine which number comes next: 1, 1, 2, 6, 24, ?

Pattern 7: Determine which number comes next: 62, 54, 57, 49, 52, ?

Pattern 8: Determine which number comes next: 4, 12, 3, 9, 0, ?

Pattern 9: Determine which number comes next: 2, 4, 8, 16, 32, ?

Pattern 10: Determine which number comes next: 43, 32, 64, 53, 106, ?

Task 3

In this task you will play 10 rounds of a decision making game. In each round you have to make several decisions. Please read the instructions carefully. You may earn a considerable sum of money, depending on the decisions you make.

In the end, one of the 10 rounds will be randomly selected and you will be randomly matched with another (anonymous) participant. You will never be matched with the same participant as in Task 1. Neither of you will know the identity of the other. The money you earn in this particular round will be paid to you at the end of the experiment in addition to your earnings from the other tasks and a fixed amount for participation. The conversion rate is set such that 100 points in the experiment correspond to $1.00 (i.e., $0.01 per point in the experiment).

Instructions

Before every round, we will roll a 6-sided die. The other participant will not be informed about the outcome of the die roll. However, he or she will be told that you have been informed the outcome of the die roll. In every round, you will be asked to send a message to the other participant. The message will correspond to a number from 1 to 6. There are six possible messages:

- Message 1: "The outcome from the roll of the 6-sided die is 1."
- Message 2: "The outcome from the roll of the 6-sided die is 2."
- Message 3: "The outcome from the roll of the 6-sided die is 3."
- Message 4: "The outcome from the roll of the 6-sided die is 4."
- Message 5: "The outcome from the roll of the 6-sided die is 5."
- Message 6: "The outcome from the roll of the 6-sided die is 6."

Your message will be shown to the other participant, and then he or she will be asked to choose a number between 1 and 6.

The choice of the number by the other participant will determine the payments in the experiment. For example:

- If he or she chooses the actual outcome of the roll of the die, then you will receive 35 points and he or she will receive 25 points.
- If he or she chooses a number different than the actual outcome, you will receive 28 points and he or she will receive 32 points.

Note that the monetary payments will change from round to round.

Again, only you will be informed of the particular monetary value connected to each number. The other participant will not be informed of these monetary values. However, he or she will be told that you have been informed of the monetary value connected to each number.

In addition, you will always see a list of different participants with which you might be matched. You will know exactly how often these participants will choose the number that you send with your message. You will have to indicate the message that you would like to send to each possible counterpart.

For your convenience, these instructions will remain available to you on all subsequent screens of this task.

Round [1-10] of 10

Before starting this round, we have rolled a 6-sided die, and obtained the outcome 3. We ask you to send a message to all possible counterparts. This message is the only information he or she will receive. He or she will not be informed about the monetary consequences of each option. In this round the monetary consequences are as follows:

- If he or she chooses the actual outcome of the roll of the die, then you will receive [Sender Low Payoff] points and he or she will receive [Receiver High Payoff] points.

- If he or she chooses a number different than the actual outcome of the roll of the die, you will receive [Sender High Payoff] points and he or she will receive [Receiver Low Payoff] points.

Below you will find a list of the 21 participants with whom you might be matched. In addition there is information how each of them will play.

Please indicate for each participant which message you would like to send him or her. You will find the six possible messages which correspond to the six numbers in the instructions below. In the end one participant will be randomly selected and you will be paid according to your decisions.

Participant 1 will choose the number that you send in your message with 0% probability. Which message would you send Participant 1?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 2 will choose the number that you send in your message with 5% probability. Which message would you send Participant 2?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 3 will choose the number that you send in your message with 10% probability. Which message would you send Participant 3?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 4 will choose the number that you send in your message with 15% probability. Which message would you send Participant 4?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 5 will choose the number that you send in your message with 20% probability. Which message would you send Participant 5?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 6 will choose the number that you send in your message with 25% probability. Which message would you send Participant 6?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 7 will choose the number that you send in your message with 30% probability. Which message would you send Participant 7?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 8 will choose the number that you send in your message with 35% probability. Which message would you send Participant 8?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 9 will choose the number that you send in your message with 40% probability. Which message would you send Participant 9?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 10 will choose the number that you send in your message with 45% probability. Which message would you send Participant 10?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 11 will choose the number that you send in your message with 50% probability. Which message would you send Participant 11?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 12 will choose the number that you send in your message with 55% probability. Which message would you send Participant 12?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 13 will choose the number that you send in your message with 60% probability. Which message would you send Participant 13?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 14 will choose the number that you send in your message with 65% probability. Which message would you send Participant 14?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 15 will choose the number that you send in your message with 70% probability. Which message would you send Participant 15?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 16 will choose the number that you send in your message with 75% probability. Which message would you send Participant 16?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 17 will choose the number that you send in your message with 80% probability. Which message would you send Participant 17?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 18 will choose the number that you send in your message with 85% probability. Which message would you send Participant 18?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 19 will choose the number that you send in your message with 90% probability. Which message would you send Participant 19?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 20 will choose the number that you send in your message with 95% probability. Which message would you send Participant 20?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Participant 21 will choose the number that you send in your message with 100% probability. Which message would you send Participant 21?

• Message 1 • Message 2 • Message 3 • Message 4 • Message 5 • Message 6

Questionnaire

Question 1: Please indicate your gender:

• Female • Male

Question 2: Please indicate your age:

Question 3: What is the highest degree/level of education you have completed?

• No schooling • Elementary school • Middle school • High school • Undergraduate school • Graduate school

Question 4: What is your average yearly household income (in USD)?

• < 20‘000 • 20‘001 - 40‘000 • 40‘001 - 60‘000 • 60‘001 - 80‘000 • 80‘001 - 100‘000 • > 100‘000 • I do not want to say

Question 5: To which religion do you belong?

• Christianity • Islam • Hinduism • Buddhism • Judaism • Folk religion • No religion • Other

Question 6: Suppose the outcome from the roll of the 6-sided die is 5. To what extent do you judge the decision to send something other than Message 5: "The outcome from the roll of the 6-sided die is 5." as:[51]

- 1: Very dishonest ... 7: Very honest
- 1: Very manipulative ... 7: Not at all manipulative
- 1: Strongly associated with personal gains ... 7: Not at all associated with personal gains
- 1: Strongly hurting the other participant ... 7: Not at all hurting the other participant

[51] The measure Prosoc is computed based on subquestions 2 and 4 while 1 and 3 serve as consistency check.

Question 7: Consider the choices of the player who sent a message. To what extent do you judge the decision to send something other than the truthful message as:[52]

- 1: Very immoral ... 7: Very moral
- 1: Not at all praiseworthy ... 7: Very praiseworthy
- 1: Not at all blameworthy ... 7: Very blameworthy
- 1: Not at all outrageous ... 7: Very outrageous
- 1: Not at all acceptable ... 7: Very acceptable

Question 8: Individuals have an opportunity to choose which information they provide to their audience, and they can modify information they obtain before passing this information on to others. Some view such modification as a violation of truthfulness; others regard it as acceptable protection of personal interests. What do you think about the value of truthfulness in such a situation? Truthfulness is about something...

- ... that one should not sacrifice, no matter what the (material or other) benefits. [1: Strongly Disagree ... 7: Strongly Agree]
- ... for which I think it is right to make a cost-benefit analysis. [1: Strongly Disagree ... 7: Strongly Agree]
- ... that cannot be measured in monetary terms. [1: Strongly Disagree ... 7:
- ... about which I can be flexible if the situation demands it. [1: Strongly Disagree ... 7: Strongly Agree]

Thank you!

You have completed the experiment session. Thank you for doing our study!

Very soon you will receive the fixed fee of $0.25. In the next hours we will randomly match you with a player for Task 1 and another player for Task 3 and also pay the bonus that you have earned depending on the rounds selected by the lotteries as well

[52] In Questions 7 and 8, I restate some variables such that a large number always corresponds to strong protected values for honesty.

as the other participants' choices. In addition you will receive the money you have earned in Task 2. We usually manage to pay participants within 24 hours.

Do you have any comments or feedback regarding this experiment?

• Yes • No

1.D.2.2 Receiver

Introduction

In this study you will play a decision making experiment that includes three tasks. Please read the instructions carefully. You may earn a considerable sum of money, depending on the decisions you make in the experiment.

In this experiment you will earn experimental points which will be converted to real money at the end. The conversion rate is set such that 100 points in the experiment correspond to $1.00 (i.e., $0.01 per point in the experiment). The money you earn in these tasks will be paid to you at the end of the experiment in addition to a fixed amount of $0.25 for participation.

Keep in mind that you will only be paid if you fully complete the study until the end.

Task 1

Instructions

In this task you will play a short decision making game. Please read the instructions carefully. You may earn a considerable sum of money, depending on the decisions you make in the experiment.

You will be randomly matched with another (anonymous) participant in this task. Neither of you will know the identity of the other.

We have rolled a 6-sided die, and told the outcome of it to the other participant, but we are not going to tell it to you.

After being informed of the roll of the die, the other participant will send a message to you. The message corresponds to a number from 1 to 6. There are six possible messages:

- Message 1: "The outcome from the roll of the 6-sided die is 1."
- Message 2: "The outcome from the roll of the 6-sided die is 2."
- Message 3: "The outcome from the roll of the 6-sided die is 3."

- Message 4: "The outcome from the roll of the 6-sided die is 4."
- Message 5: "The outcome from the roll of the 6-sided die is 5."
- Message 6: "The outcome from the roll of the 6-sided die is 6."

We will ask you to choose a number between 1 and 6. The message you receive is the only information you will have regarding the roll of the die. Your choice of a number will determine the payments according to two different options, known only to the other participant.

If you will choose the same number as the number that came up in the roll of the die you will be paid more than if the number is different than the actual number. For your convenience, these instructions will remain available to you on all subsequent screens of this task.

Decisions

The other participant will send you a message. For every possible message we ask you to choose one of the six numbers. Your choice will determine the payments to both of you.

Decision 1: Suppose the other participant sent you: Message 1: "The outcome from the roll of the 6-sided die is 1."

Given this message the number I choose is:

• 1 • 2 • 3 • 4 • 5 • 6

Decision 2: Suppose the other participant sent you: Message 2: "The outcome from the roll of the 6-sided die is 2."

Given this message the number I choose is:

• 1 • 2 • 3 • 4 • 5 • 6

Decision 3: Suppose the other participant sent you: Message 3: "The outcome from the roll of the 6-sided die is 3."

Given this message the number I choose is:

• 1 • 2 • 3 • 4 • 5 • 6

Decision 4: Suppose the other participant sent you: Message 4: "The outcome from the roll of the 6-sided die is 4."

Given this message the number I choose is:

• 1 • 2 • 3 • 4 • 5 • 6

Decision 5: Suppose the other participant sent you: Message 5: "The outcome from the roll of the 6-sided die is 5."

Given this message the number I choose is:

• 1 • 2 • 3 • 4 • 5 • 6

Decision 6: Suppose the other participant sent you: Message 6: "The outcome from the roll of the 6-sided die is 6."

Given this message the number I choose is:

• 1 • 2 • 3 • 4 • 5 • 6

Question regarding your decisions

Out of 100 possible counterparts, how many do you think will send you a truthful message (i.e., send the message with the number of the actual outcome from the roll of the 6-sided die)?

Task 2

In this task you will have to complete numerical patterns. You will see 10 patterns and your task will be to determine the number that comes next in each series.

For each correctly solved pattern you will earn 10 points towards your total payment at the end of the experiment. The conversion rate is set such that 100 points in the experiment correspond to $1.00 (i.e., $0.01 per point in the experiment).

You have 120 seconds to determine as many missing numbers as possible. Once time is over you will be automatically redirected to the next page. You may start the clock by clicking on the button below.

This task is followed by some general questions.

Task 2

Time left to complete this page: [mm:ss]

Pattern 1: Determine which number comes next: 13, 17, 19, 23, 29, ?

Pattern 2: Determine which number comes next: 1, 4, 9, 16, 25, ?

Pattern 3: Determine which number comes next: 3, 18, 9, 54, 27, ?

Pattern 4: Determine which number comes next: 3, 4, 8, 11, 44, ?

Pattern 5: Determine which number comes next: 2, 5, 10, 17, 28, ?

Pattern 6: Determine which number comes next: 1, 1, 2, 6, 24, ?

Pattern 7: Determine which number comes next: 62, 54, 57, 49, 52, ?

Pattern 8: Determine which number comes next: 4, 12, 3, 9, 0, ?

Pattern 9: Determine which number comes next: 2, 4, 8, 16, 32, ?

Pattern 10: Determine which number comes next: 43, 32, 64, 53, 106, ?

Task 3

Instructions

Please read the instructions carefully. You may earn a considerable sum of money, depending on the decisions you make in this task.

In this task you will face decision situations with 10 anonymous participants. Neither of you will know the identity of the other.

We have rolled a 6-sided die, and told the outcome of it to the other participant, but we are not going to tell it to you.

After being informed of the roll of the die, the other participant will send a message to you. The message corresponds to a number from 1 to 6. There are six possible messages:

- Message 1: "The outcome from the roll of the 6-sided die is 1."
- Message 2: "The outcome from the roll of the 6-sided die is 2."
- Message 3: "The outcome from the roll of the 6-sided die is 3."
- Message 4: "The outcome from the roll of the 6-sided die is 4."
- Message 5: "The outcome from the roll of the 6-sided die is 5."
- Message 6: "The outcome from the roll of the 6-sided die is 6."

The message you receive is the only information you will have regarding the roll of the die. We will ask you to choose a number between 1 and 6. In addition, you are forced to choose the number that was stated in the message exactly [ ] times in total (out of 60 decisions).

In the end a lottery will randomly determine with which of the seven participants you will be matched. Your choice of a number will determine the payments in the experiment according to two different options, known only to the other participant. If you will choose the same number as the number that came up in the roll of the die you will be paid more than if the number is different than the actual number.

For your convenience, these instructions will remain available to you on all subsequent screens of this task.

Decisions for Participant [1-10]

Participant [1-10] will send you a message. For every possible message we ask you to choose one of the six numbers. Your choice will determine the payments to both of you.

Keep in mind that you must choose the same number as the one in the message in exactly [ ] out of the 60 cases in total.

You have already chosen a number equal to the number stated in the message [ ] times. This means that you must do so in [ ] more cases.

Decision 1: Suppose the other participant sent you: Message 1: "The outcome from the roll of the 6-sided die is 1."

Given this message the number I choose is:

• 1 • 2 • 3 • 4 • 5 • 6

Decision 2: Suppose the other participant sent you: Message 2: "The outcome from the roll of the 6-sided die is 2."

Given this message the number I choose is:

• 1 • 2 • 3 • 4 • 5 • 6

Decision 3: Suppose the other participant sent you: Message 3: "The outcome from the roll of the 6-sided die is 3."

Given this message the number I choose is:

• 1 • 2 • 3 • 4 • 5 • 6

Decision 4: Suppose the other participant sent you: Message 4: "The outcome from the roll of the 6-sided die is 4."

Given this message the number I choose is:

• 1 • 2 • 3 • 4 • 5 • 6

Decision 5: Suppose the other participant sent you: Message 5: "The outcome from the roll of the 6-sided die is 5."

Given this message the number I choose is:

• 1 • 2 • 3 • 4 • 5 • 6

Decision 6: Suppose the other participant sent you: Message 6: "The outcome from the roll of the 6-sided die is 6."

Given this message the number I choose is:

• 1 • 2 • 3 • 4 • 5 • 6

Questionnaire

Question 1: Please indicate your gender:

• Female • Male

Question 2: Please indicate your age:

Question 3: What is the highest degree/level of education you have completed?

• No schooling • Elementary school • Middle school • High school • Undergraduate school • Graduate school

Question 4: What is your average yearly household income (in USD)?

• < 20'000 • 20'001 - 40'000 • 40'001 - 60'000 • 60'001 - 80'000 • 80'001 - 100'000 • > 100'000 • I do not want to say

Question 5: To which religion do you belong?

• Christianity • Islam • Hinduism • Buddhism • Judaism • Folk religion • No religion • Other

Question 6: Suppose the outcome from the roll of the 6-sided die is 5. To what extent do you judge the decision to send something other than Message 5: "The outcome from the roll of the 6-sided die is 5." as:

- 1: Very dishonest ... 7: Very honest
- 1: Very manipulative ... 7: Not at all manipulative
- 1: Strongly associated with personal gains ... 7: Not at all associated with personal gains
- 1: Strongly hurting the other participant ... 7: Not at all hurting the other participant

Question 7: Consider the choices of the player who sent you a message. To what extent do you judge the decision to send something other than the truthful message as:

- 1: Very immoral ... 7: Very moral
- 1: Not at all praiseworthy ... 7: Very praiseworthy
- 1: Not at all blameworthy ... 7: Very blameworthy
- 1: Not at all outrageous ... 7: Very outrageous
- 1: Not at all acceptable ... 7: Very acceptable

Question 8: Individuals have an opportunity to choose which information they provide to their audience, and they can modify information they obtain before passing this information on to others. Some view such modification as a violation of truthfulness; others regard it as acceptable protection of personal interests. What do you think about the value of truthfulness in such a situation? Truthfulness is about something...

- ... that one should not sacrifice, no matter what the (material or other) benefits. [1: Strongly Disagree ... 7: Strongly Agree]
- ... for which I think it is right to make a cost-benefit analysis. [1: Strongly Disagree ... 7: Strongly Agree]
- ... that cannot be measured in monetary terms. [1: Strongly Disagree ... 7:
- ... about which I can be flexible if the situation demands it. [1: Strongly Disagree ... 7: Strongly Agree]

Thank you!

You have completed the experiment. Thank you for doing our study!

Very soon you will receive the fixed fee of $0.25. In the next hours we will randomly match you with a player for Task 1 and another player for Task 3 and also pay the bonus that you have earned depending on the rounds selected by the lotteries as well as the other player's choice. In addition you will receive the money you have earned in Task 2. We usually manage to pay participants within 24 hours.

Do you have any comments or feedback regarding this experiment?

• Yes • No

2 Model Selection from Experimental Data: Evidence from Individual Lying Behavior*

with Joel Sobel and Alexander F. Wagner

2.1 Introduction

Economic theory designs general models to organize behavioral observations and to make predictions in novel situations. Experimental economics identifies simple situations in which standard models fail to organize data and uses the experimental evidence to support alternative assumptions. In this paper, we show that offering a new model that organizes a particular data set better than a given candidate model does not provide evidence that the new model is useful in other situations. While the basic notion of a difference between in-sample and out-of-sample performance is well-known, its relevance for experimental research may be underappreciated and undocumented.

This paper collects experimental data on two familiar games of communication that we adapt from Gneezy (2005) and Erat and Gneezy (2012). After the Sender privately learns the state, she is allowed to send a message about the state to the Receiver before the latter takes an action that determines the payoff of both players. We specify monetary payoffs such that the Sender would always like the Receiver not to match the true state, thus providing incentives to lie. We say that the Sender lies if her statement does not match her private information. We run variations of these experiments with a heterogeneous sample of Amazon Mechanical Turk (mTurk) users in the US, collecting data on 9000 decisions of 400 individuals. Consistent with existing studies, we find that some Senders tell the truth even though doing so does not maximize their monetary payoff.

We use a subset of the data to classify subjects into five types. We then study how well this classification organizes the remaining data. The analysis provides insight into the range of preferences for honesty present in the population. It also delivers more general

* We would like to thank Johannes Abeler, Zachary Breig, Helga Fehr-Duda, Michel Maréchal, Bettina Rockenbach, participants at the Spring School in Behavioral Economics at UCSD, the Morality, Incentives and Unethical Behavior Conference, the UCSD Economics Department brown bag seminar, and the UZH PhD Poster Workshop for very helpful comments and suggestions. Part of this paper was developed while Bögli was a visiting PhD student at UCSD. Financial support from the Swiss Finance Institute is gratefully acknowledged.

insights into the problem of model selection and into how model specifications derived from one data set lead to useful predictions in novel contexts. We hope to stimulate further work aimed at comparing how to select one model from several available ones.

The specific games we study are of particular interest to behavioral economics. There are many situations in which people can earn direct rewards from dishonest behavior. While people take advantage of these opportunities, there is substantial evidence that at least some people are reluctant to do so. Understanding the extent to which a narrow definition of utility (maximizing monetary behavior) fails to describe behavior is a necessary first step to understanding how to build and maintain institutions that lead to desirable outcomes even in the presence of temptations to lie.

A convincing body of evidence indicates that people are willing to lie when lying brings direct economic benefits, but that some people also do not like to lie. We wish to organize our findings by trying to set the extent to which subjects' behavior can be described by a small set of pre-specified utility functions. Based on Abeler et al.'s (2016) comprehensive meta-study on lying we consider five broad categories of preferences: (i) standard economic theory, (ii) inequity aversion, (iii) reference-dependent preferences, (iv) lying costs, and (v) reputation or social norms.

Subjects in the first category simply maximize their expected monetary payoff. These subjects lie whenever there are economic benefits to lying. Inequity averse agents dislike differences in payoffs between themselves and the other player.[1] Inequity-averse individuals prefer to behave honestly when reporting dishonestly generates a large difference in payoffs compared to the status quo. Reference-dependent preferences posit that utility depends on personal expectations. These individuals derive utility in relation to a reference point, e.g., the expected payoff.[2] Lying costs correspond to the idea that participants incur a private cost when telling a lie. Moral reasons, self-image concerns or injunctive social norms may lead to such costs.[3] Finally, reputation and descriptive social norms might be associated with honesty.[4] A higher fraction of individuals acting honestly may increase the utility costs of dishonesty, and participants may dislike to act in the same way as dishonest people.

[1] Examples of models of inequity aversion are Fehr and Schmidt (1999) or Bolton and Ockenfels (2000).

[2] Köszegi and Rabin (2006) provide the seminal model featuring expectations as reference points.

[3] There are several different lying cost characterizations in the literature. E.g., Kartik (2009), López-Pérez and Spiegelman (2013), or Gneezy et al. (2016)

[4] Examples of studies that consider some form of a reputation for honesty, social identity or social norms include: Akerlof (1983), Weibull and Villa (2005), Mazar et al. (2008), and Fischbacher and Heusi (2013) among others. Note that experiments such as Fischbacher and Heusi (2013) are anonymous. They conclude that reputational concerns towards others are not critical. Ellingsen and Johannesson (2008), too, conclude that social esteem may, in fact, play a role even in anonymous settings. Nonetheless explicit reputation and repeated interaction could likely induce an additional tendency to behave more or less honestly.

Although some evidence exists on the empirics of truthfulness preferences,[5] little is known about the heterogeneity of preferences within the population and the relative importance of different preferences in explaining lying behavior. Knowing the distribution of preferences for truthfulness would improve our understanding of human lying behavior and possibly allow us to anticipate different truth-telling patterns. A similar situation arises when researchers try to select the best model of behavior in settings involving potential pro-social choices. Methodologically, our contribution, therefore, is to provide a framework that can also be employed in other instances where researchers are interested in assessing plausible preference specifications.

First, we estimate parameters of a simple utility specification conditional on being a member of either type for each participant. Then, we let the data tell us which of all available types explains an individual's behavior best. As we consider all possible combinations of types, we have to account for 31 possible worlds (5 one-type worlds, 10 two-type worlds, 10 three-type worlds, 5 four-type worlds, and 1 five-type world). Hence in a given world not every type is available. We assign each participant to the type among all available types of a given world that explains individual behavior best.[6]

The second step of our model selection analysis is to analyze how well the 31 worlds fit the data *in-sample* and *out-of-sample*. We split the sample into an estimation part (in-sample) and a prediction part (out-of-sample). As the in-sample part we define the set of decisions which is used for the first step of our analysis, i.e., estimation and classification. The out-of-sample part consists of all remaining decisions. We derive the parameter estimates and the classification based on the in-sample part and analyze how well these parameters fit behavior out-of-sample. In addition, we are interested in how the quality of the in-sample and the out-of-sample fit depend on the number of decisions used to estimate and classify participants into types.

We compute a quality measure in order to assess the performance of the fits. We compute the utilities of lying and not lying based on the estimated parameters for every participant and take the message that yields a higher utility. We then compare predicted decisions to real decisions and compute the number of correctly fitted decisions in relation to the total number of decisions, this is our accuracy measure.

[5] Abeler et al. (2014) find lying costs to be large and widespread among the German population. Gibson et al. (2013) provide evidence for heterogeneity of these costs across and within individuals. Gneezy et al. (2013) classify people into broad types, including those who never lie, those who always lie, and those who react to economic incentives.

[6] A central feature of our analysis is that we only use behavior to classify individuals. The alternative is to run regressions of individual behavior on certain measurable characteristics of individuals and then to use the estimated regression parameters to make predictions for the future behavior of those individuals, conditional on their characteristics. We show that these characteristics (except some that are in general hard to measure, such as intrinsic values) explain little of variation in behavior.

Because the number of conceivable worlds is large, we restrict our attention to unique and well performing worlds. Specifically, we eliminate worlds *equivalent* to a world that allows for fewer types and *dominated* worlds. Two worlds with different numbers of types are considered *equivalent* if they assign participants to the same types. This implies that at least one type is redundant and contains no members. As a result, the world in which a given type is redundant can be eliminated in favor of a smaller world that considers the same types except the redundant. *Dominated* worlds are those whose in- or out-of-sample fit is strictly worse than another world's, regardless of the number of observations used to estimate parameters and to classify participants to types.

Our results are similar in the two slightly different environments: First, the quality of the in-sample fit exhibits a (broadly) decreasing convex and the out-of-sample fit an increasing concave shape with respect to the number of decisions used for the estimation. That is, adding data points allows us to fit in-sample worse at a decreasing rate and out-of-sample better at a decreasing rate. Second, worlds that fit the data well in-sample do not necessarily perform well out-of-sample and vice versa. The quality of an in-sample fit conveys no information about the quality of an out-of-sample fit for a fixed number of observations in each of the two samples. Third, the size of the in-sample part is important for determining the best possible world in-sample and out-of-sample. Finally, we provide indicative evidence that the share of each type within the whole population is quite stable across the two situations we study.

When only a small number of decisions is used to estimate and classify, lying costs explain behavior best in-sample. However, they do poorly in-sample once more data are collected. More importantly, they do poorly relative to other models in predicting lying behavior out-of-sample. A world consisting of three types – reference-dependent preferences, lying costs, and a concern for reputation and descriptive social norms – organizes lying behavior best when combining in-sample and out-of-sample performance. In this three-type world (and in the benchmark five-type world) lying costs and reputation/social norm concerns make up the largest shares in the population. The natural caveat here is that in a richer setting – for example, one that involves repeated interactions – other preference models would perhaps be available.

We reach a simple but instructive conclusion. The quality of the in-sample fit of a model does not say anything about how well that model will be able to explain future behavior. Similarly, worlds that perform comparably poorly in-sample might do a decent job in organizing future data sets. As a consequence, it is crucial that researchers are very cautious when deciding about a preferred model based on an in-sample methodology only. The most important safeguard against this error is to collect data on a large number of decisions.

Our study makes methodological and substantive contributions. Methodologically, this paper contributes to a small recent literature on model selection in experiments (e.g., Ericson et al. (2015), Peysakhovich and Naecker (2017), and Breig (2017)). These papers study model selection in the domains of preferences over time, risk, and/or ambiguity. The latter two studies provide evidence that it depends on the context whether a better in-sample fit also implies a more accurate out-of-sample prediction. All of these studies focus on one-type worlds.[7] We instead also consider worlds consisting of several types to which participants are assigned according to their behavior. This allows us to study how the distribution of types depends on the number of observations, and to document that having more types available does not necessarily improve the predictions.

Thematically, our study links to the broad experimental literature on strategic information transmission games (e.g., Dickhaut et al. (1995), Blume et al. (1998), Cai and Wang (2006), Wang et al. (2010) among others) and deception in general (Bucciol et al. (2013), Fischbacher and Heusi (2013)). However, our focus lies, on the one hand, on model selection in strategic information transmission environments per se and, on the other hand, on the heterogeneity of preferences that lead to untruthful reporting of agents. Our paper is most closely related to the meta-study of Abeler et al. (2016) as we rely on five types which correspond to different broad categories of explanations according to their structural characterization. In addition, both Abeler et al. (2016) and we aim at investigating what kind of preferences for honesty can best explain lying. While they apply a meta-study of existing data, excluding models one-by-one that cannot organize these in-samples, we collect new data, which allows us to run an extensive statistical analysis. Most importantly, we are able to compare how well different models fit lying data in-sample and out-of-sample.

The remainder of the paper is organized as follows. Section 2.2 explains the experimental design. In Section 2.3, we introduce the estimation as well as the classification and highlight the statistical methodology. Section 2.4 describes the sample as well as its descriptive statistics. Section 2.5 conducts the model selection and provides results. Section 2.6 is a conclusion.

[7] Only Breig (2017) in one analysis investigates the performance of a world that allows for two different risk preferences.

2.2 Experimental Methodology

2.2.1 Games

We slightly adapt two well-established sender-receiver games from the literature. Although both share the conceptual roots in the cheap-talk literature, their structures differ somewhat.

Game A follows Gneezy (2005) where the Sender is informed about two payoff options (named Option X and Y) conditional on the Receiver's action. Knowing the payoff matrix, she must provide a message concerning the payoffs to the Receiver. The message space consists of two messages:

1. "Option X will earn you more money than Option Y."

2. "Option Y will earn you more money than Option X."

The Receiver does not (and will never) know the payoffs. His choice (either Option X or Y) determines the payment to both players. Finally, the players are paid according to the Receiver's action.

In **Game B** we adapt the game of Erat and Gneezy (2012). Nature rolls a 6-sided die in the beginning. The Sender privately observes the outcome from the roll of the 6-sided die. In addition, the Sender is informed about the two payoff options: Option X if the Receiver chooses the actual number from the roll of the die, and Option Y if the Receiver chooses any number different from the actual outcome of the roll of the 6-sided die. The Sender must send a message about the actual outcome to the Receiver. The message space consists of six messages. For $i = \{1, 2, \ldots 6\}$, message i is:

> The outcome from the roll of the 6-sided die is i.

The Receiver does not (and will never) know the payoffs. He chooses one number upon receiving the Sender's message. The payments to both players depend on whether the Receiver picked the actual outcome of the roll of the die.

2.2.2 Payoffs

We use the same set of payoff matrices for both games. We specify the payoffs in order to provide economic incentives to lie ($\Pi^S(a, \theta) > \Pi^S(\theta, \theta)$ when $a \neq \theta$), i.e., the Sender's payoff Π^S if the Receiver chooses any action a that is different from the state θ is larger than the Sender's payoff if the Receiver's action matches the state. We further assume that monetary payoffs depend only on whether $a = \theta$ so that the quantity $\Pi^S(a, \theta) - \Pi^S(\theta, \theta)$ is constant when $a \neq \theta$. When $a \neq \theta$, this quantity is the monetary cost of stating the truth, which we denote by $\Xi > 0$ and call ECOST. In both games, the Sender's payoff is maximized when the Receiver chooses the option that does not lead to the Receiver's higher payoff. Assuming that the Receiver chooses the outcome that maximizes his payoffs under the assumption that the Sender's message is accurate,[8] the Sender has monetary incentives to lie. We vary Ξ to investigate the consequences of changing ECOST on truth-telling behavior. Overall, we use 15 differing values of ECOST ranging from \$0.05 to \$0.75.[9]

Table 2.A.1 in Appendix 2.A displays all payoff matrices. As an illustration, one specific payoff matrix is displayed in Figure 2.1. This particular payoff matrix corresponds to the matrix depicted in the first row of the $\Xi = \$0.25$ box in the first column of Table 2.A.1. While participants see the payoffs in experimental currency units that are exchanged at a rate of \$0.05 per unit, we directly state payoffs in terms of \$ values here.

	Receiver	
	$a = \theta$	$a \neq \theta$
Sender	\$1.15 \$1.10	\$0.60 \$1.35

Figure 2.1: Payoff matrix example

Given this, the payoffs are $(\$1.10, \$1.15)$ if $a = \theta$, i.e., \$1.10 for the Sender and \$1.15 for the Receiver, and $(\$1.35, \$0.60)$ if $a \neq \theta$, i.e., \$1.35 for the Sender and \$0.60 for the Receiver. Consequently, the monetary incentives of lying amount to \$0.25 (= \$1.35 – \$1.10).

Each participant plays one game either in the role of a Sender or a Receiver. We collect comprehensive data for Senders. Specifically, we put the Sender in $K = 45$ differing and randomly ordered decision situations. We apply a direct elicitation method for the Sender's strategies. Each of the 15 ECOST levels is played three times. While the Sender plays 45 independent rounds with varying payoff matrices, we use the strategy-method to elicit the Receiver's behavior as he has no knowledge about the payoff options and hence his strategic decision situation does not vary across rounds.

In addition, we elicit the player's first-order beliefs about the other player's action. Finally, a questionnaire eliciting demographic characteristics and protected values (as in Tanner et al. (2009) and Gibson et al. (2013)) and prosocial concerns completes our

[8] This assumption is consistent with the data.

[9] The implied effect on the Sender's average compensation is between 3.1% (= 0.05/1.60) and 46.9% (= 0.75/1.60).

experiment.[10] In order to determine the payments to both players, one of the Sender's 45 rounds is randomly selected and matched with the Receiver's corresponding action. The anonymously matched participants are paid according to the outcome induced by their behavior. All participants were paid within 24 hours after completion of the experiment.

We ran two sessions of the experiment in March and August 2015. The experiment is programmed and conducted with the experimental software oTree (Chen et al. (2015)). Appendix 2.B contains the complete set of experimental instructions.

2.3 Estimation, Classification and Model Selection

2.3.1 Types

Let the Sender's utility in a given state conditional on belonging to type t be

$$U_i^t(a, \theta, m_i) = \Pi^S(a, \theta) - \Psi_i^t(a, \theta, m_i),$$

where $\Pi^S(\cdot)$ denotes the Sender's payoff if nature draws θ, the Sender's message is m_i, and the Receiver chooses a. Varying the functional form of $\Psi_i^t(\cdot)$ determines the type of model that describes the Sender's preferences.[11] The literature of deception and lying suggests several models to describe lying behavior. Abeler et al. (2016) provide a comprehensive meta-study. We follow their general set up and restrict attention to five broad categories of lying costs: (i) standard economic theory (EC), (ii) social preferences in the form of inequity aversion (IA), (iii) reference-dependent preferences (RD), (iv) lying costs (LC), and (v) reputation or social norms (SN).[12]

(i) The first type of agent maximizes her own monetary payoff according to the standard economic model (EC). Consequently, EC participants are willing to lie whenever there are economic benefits to lying as $\Psi_i^{\text{EC}}(\cdot) \equiv 0$. Since our payoff matrices always feature such incentives to tell a lie, participants are expected to report dishonestly if they believe the Receiver to follow their message in more than one half of all cases in

[10] As prosociality is not our main focus, we use a simple prosocial concerns measure that is based on two survey questions. There are more sophisticated prosocial concerns measures (see, e.g., Van Lange et al. (1997)).

[11] Note that Ψ_i^t may sometimes depend on other parameters.

[12] Of course these five types do not form a complete taxonomy. While we stick to models suggested by Abeler et al. (2016), there might be other models that could do a much better job in organizing data. In addition, even for the broad types that we consider, there are other specifications how such preferences could be characterized. Our approach therefore serves as a proof of concept for a popular selection of models rather than a conclusive analysis on the whole universe of possible lying models.

Game A and more than one sixth of all cases in Game B.[13] EC Participants derive utility $U_i^{\mathrm{EC}}(a,\theta,m_i) = \Pi^S(a,\theta)$.

(ii) Interdependent preferences imply that utility depends on other participants' payoffs. Specifically, we consider inequity aversion where participants might have a distaste for a big difference in the payoffs between Sender and Receiver. Bolton and Ockenfels (2000) and Fehr and Schmidt (1999) introduce models of inequity aversion (IA). Individuals motivated by IA would hence prefer to behave honestly if reporting untruthfully implied a substantial difference in payoffs compared to the status quo. We consequently define $\Psi_i^{\mathrm{IA}}(\cdot) \equiv \max\Big(\Pi^R(a,\theta) - \Pi^S(a,\theta), 0\Big)$ and consider the utility specification

$$U_i^{\mathrm{IA}}(a,\theta,m_i) = \Pi^S(a,\theta) - \gamma_{IA}\max\Big(\Pi^R(a,\theta) - \Pi^S(a,\theta), 0\Big),$$

where γ_{IA} denotes the coefficient that loads on to the difference between the Receiver's and the Sender's payoff. From a Sender's perspective this difference corresponds to the negative inequity in a given state. Due to the design of our payoff matrices negative and positive inequity are perfectly (negatively) correlated. As a consequence, we cannot differentiate between an increase in negative inequity and a decrease in positive inequity or vice versa. Hence we only consider negative inequity.

(iii) Reference-dependent (RD) preferences suggest that utility depends on personal expectations. RD individuals derive utility in relation to a reference point, e.g., the expected payoff. Köszegi and Rabin (2006) provide the seminal model featuring expectations-dependent reference points. We assume $\Psi_i^{\mathrm{RD}}(\cdot) \equiv \Big(\Pi^S(a,\theta) - \mathbb{E}\Big[\Pi^S(a,\theta)\Big]\Big)$ and compute the utility specification as

$$U_i^{\mathrm{RD}}(a,\theta,m_i) = \Pi^S(a,\theta) + \gamma_{RD}\Big(\Pi^S(a,\theta) - \mathbb{E}\Big[\Pi^S(a,\theta)\Big]\Big),$$

where γ_{RD} reflects a coefficient that measures the importance of the reference point. The expectation of the payoff depends on the Sender's first-order belief, i.e., the subjective likelihood that the Receiver's action will match the message. If the possible gain of lying is high compared to the expected payoff in the status quo, participants will tend to lie more (Garbarino et al. (2017)).

[13] In Game A, if the Receiver follows the Sender's message with probability p, the gain from lying is $(2p-1)\Xi$. In Game B, if the Receiver follows the Sender's message with probability p and otherwise selects randomly, the Sender obtains her preferred action with probability $p + .8(1-p)$ if she lies and with probability $(1-p)$ if she tells the truth. Hence her monetary payoff is maximized by lying if $p > 1/6$. In Appendix 2.C, we illustrate the expected utility computation for all types.

(iv) Lying costs (LC) refer to the idea that participants incur a private cost when telling a lie. Moral reasons, self-image concerns or injunctive social norms may lead to such costs. There are several different lying cost characterizations in the literature: Pure lying aversion postulates that lies are costly since uttering a lie per se triggers bad feelings implying losses in utility (e.g., Kartik (2009), López-Pérez and Spiegelman (2013)). Others, e.g., Gibson et al. (2013), suggest that lying costs are increasing in the incentives provided to lying, or, e.g., Gneezy et al. (2016) argue that they depend on the three dimensions: payoff of the player, outcome of the private information, and likelihood of the lie.

For convenience we posit lying costs to be convex and increasing in the monetary incentives of lying and define $\Psi_i^{\mathrm{LC}}(\cdot) \equiv (\alpha + \beta\Xi + \rho\Xi^2)\, \mathbb{1}_{m_i \neq \theta}$. Whenever these lying costs are smaller than the possible gain of deviating from the truth, LC individuals will report untruthfully. The utility representation reads

$$U_i^{\mathrm{LC}}(a, \theta, m_i) = \Pi^S(a, \theta) - \left(\alpha + \beta\Xi + \rho\Xi^2\right) \mathbb{1}_{m_i \neq \theta},$$

where α, β, and ρ are the coefficients for the constant, the linear and the convex term of ECOST, respectively. Lying costs occur if the Sender lies, i.e., if the message deviates from the state.

(v) Social norms (SN) accounts for the effects of reputation and descriptive social norms that might be associated with honesty. The idea of reputation for honesty is that participants dislike to act in the same way as dishonest people. Since only the experimenter observes whether a participant reports honestly, reputation for honesty in our setting does not consider how participants are perceived by others. Similarly, descriptive social norms for honesty increase the utility costs of dishonesty. We assume $\Psi_i^{\mathrm{SN}}(\cdot) \equiv \sum_{j \neq i} \frac{\mathbb{1}_{m_j = \theta}}{N} \mathbb{1}_{m_i \neq \theta}$ and use the utility characterization

$$U_i^{\mathrm{SN}}(a, \theta, m) = \Pi^S(a, \theta) - \gamma_{SN} \sum_{j \neq i} \frac{\mathbb{1}_{m_j = \theta}}{N} \mathbb{1}_{m_i \neq \theta},$$

where γ_{SN} is the individual coefficient that loads on the average proportion of truthful reports in a given situation. The higher the proportion of truthful reports in a given game the lower are the incentives to lie due to concerns for reputation and social norms. Consequently, SN participants are expected to report deceitfully more often if other people do so frequently too. Unlike the other specification, in this specification m represents the entire set of messages sent by the population, with m_j the message of Sender j. As only the experimenter observes whether a participant reports honestly, reputation for honesty in our setting does not consider how participants are perceived by others. As we analyze several types in one single data set our simplistic form of SN should be sufficient. However, there are also other characterizations of reputation and social norms. Studies that consider some form of a reputation for honesty or social norms include, among others, Akerlof (1983), Weibull and Villa (2005), Mazar et al. (2008), Fischbacher and Heusi (2013), López-Pérez and Spiegelman (2013), Dufwenberg and Dufwenberg (2016), Khalmetski and Sliwka (2017). Mazar et al. (2008) and Fischbacher and Heusi (2013) argue that even in anonymous settings such motivations can be at play.

2.3.2 Estimation

In order to estimate the parameters we maximize the log-likelihood function given by

$$L_i^t = \sum_{k=1}^{K} \log(l_{ik}^t),$$
$$l_{ik}^t = \left(\frac{e^{\eta(i,k|t)}}{1+e^{\eta(i,k|t)}}\right)^{1-T_{ik}} \left(\frac{1}{1+e^{\eta(i,k|t)}}\right)^{T_{ik}},$$

where $\eta(i,k \mid t) = \mathbb{E}[U_i^t(a_k,\theta_k,\theta_k') - U_i^t(a_k,\theta_k,\theta_k)]/\lambda_i,\ \ \theta_k' \neq \theta_k$ is the gain in expected utility associated with lying if Sender i has type $t \in \{\text{EC, IA, RD, LC, SN}\}$ preferences.

In Appendix 2.C, we provide the respective expected utility computations conditional on belonging to either of the five types. The parameter λ_i reflects a scale parameter measuring the degree of erroneous play. We allow for players making suboptimal decisions and maximizing $U_i^t(\cdot)+\lambda_i\epsilon$ instead, where ϵ denotes logistic errors. T_{ik} is a binary variable taking the value one if the Sender i states the truth in decision k, and zero else.

Because we are only interested in Sender behavior and how incentives rather than first-order beliefs influence behavior, we neglect Receiver behavior and assume Senders' first-order beliefs to equal one, i.e, Receivers to be credulous. As a consequence, our analysis is not subject to possible belief misstatements by participants who want to look more ethical and thus adjust their self-reported beliefs as shown by, e.g., Andreoni and Sanchez (2014). However, we also conduct the full analysis using self-reported first-order beliefs to derive expected utilities. We find qualitatively similar results, which are available on request.

We apply a maximum likelihood procedure to estimate this random effects logit model.[14] We estimate the parameters in $\Psi(\cdot)$ and the degree of erroneous play at the in-

[14] For computational simplicity we set lower and upper bounds for the parameter estimates. Neither a change in the bounds nor the initial values of the parameters has a substantial effect on this paper's findings.

dividual level conditional on the five different types. Having computed the individual likelihood for each type, we proceed with the classification of all participants.

2.3.3 Classification

Because it is our goal to compare and analyze worlds with several types, we have to first sketch out the whole space of possible type combinations. All possible type combinations yield 31 different worlds: 5 worlds with only one type, 10 worlds with two types, 10 worlds with three types, 5 worlds with four types, and one world with all five types. The list of the 31 models is displayed in Table 2.D.1 in Appendix 2.D. Having a broad variety of different worlds, we can analyze what combination of types explains the data best.

After having computed the log-likelihood conditional on belonging to each single type in a given world, we assign every participant to the type which yields the highest log-likelihood in-sample. Based on the resulting assignment we sum the respective log-likelihoods L_i^t of all participants in order to derive the total log-likelihood of a given world. Note that for this procedure we only use behavioral information. This approach deliberately leaves aside potentially relevant information that could be gleaned from other personal characteristics.

2.3.4 Model Selection

Having completed the classification, our goal is to analyze how well the 31 worlds fit the data *in-sample* as well as *out-of-sample*. First, we split the sample into an estimation part (in-sample) and a prediction part (out-of-sample). We use the in-sample decisions to estimate individual parameters of the different types and to classify people into types. The out-of-sample decisions are those not used for estimation and classification. We use the classification based on in-sample decisions to compute the quality of the out-of-sample estimates. We are interested in how the quality of the in-sample and the out-of-sample fits depend on the number of decisions used to estimate and classify participants into types.

We compare the different worlds as follows: Step 1, we split the sample into two parts according to the number of decisions k. Thus, we take $k \in \{1, 2, 3, \dots, 44\}$ decisions in the in-sample part and $45 - k$ decisions in the out-of-sample part. We start by setting $k = 1$. Step 2, we randomly select k decisions from the set of all decisions. This random selection determines which specific decisions we consider for the in-sample. Step 3, we estimate individual parameters conditional on being a member of a given type and run the classification by assigning subjects to types based on the individual conditional log-likelihoods. Step 4, we derive the *in-sample* fit, i.e., we aggregate individual log-likelihoods, conditional on k and the specific random draw and compute the implied *out-of-sample* fit based on the estimated parameters. Step 5, we repeat the steps 2 to 4 for M different random selections of the k decisions and then derive average in-sample and out-of-sample performances for all worlds across the M draws. Finally, we set $k_{new} = k + 1$ and repeat all the previous steps.[15]

The quality measure we use to assess in-sample and out-of-sample fits is the accuracy with respect to the share of correctly fitted decisions. We compute the expected utilities of lying and not lying based on the estimated parameters for every participant and predict that the participant will choose the message that yields a higher utility. We then compare the predicted decisions to real decisions and compute the number of correctly fitted decisions in relation to the total number of decisions; this is our overall accuracy measure. An alternative quality measure is the normalized log-likelihood, which can be computed directly from individual log-likelihood values normalized by the number of decisions considered in the sample. However, the normalized log-likelihood only behaves sensibly as a way to compare in-sample performance. The measure explodes when assessing out-of-sample performance as its value changes dramatically if there is a single observation for one individual that is unlikely to be derived from the model under consideration. Because the out-of-sample fit is not the result of an optimization, this happens frequently.

2.4 Data and Descriptive Evidence

2.4.1 Data Collection and Sample Characteristics

We recruited 400 US participants (100 Senders and 100 Receivers per game) on mTurk. We required subjects to have successfully completed at least 1000 HITs (human intelligence tasks on mTurk) with an overall approval rate of more than 95%.[16] For each participant, we included two understanding questions in the experiment. The first understanding was adequately solved by 77% of all participants. 94% were able to answer the second question correctly. This suggests that participants learned through the feedback to the first question and that they achieved an overall satisfactory level of understanding.

Our sample consists of 53% males and 47% females. Compared to laboratory studies our sample is very heterogeneous featuring an age range from 16 to 76 years and a va-

[15] Our approach is consistent with the idea of cross-validation, an unbiased method to evaluate models with different numbers of parameters (e.g., Kohavi (1995), Arlot and Celisse (2010)).

[16] There are several studies showing that experiments executed on mTurk yield results consistent with these obtained in common laboratory settings even though the stakes are usually smaller (e.g., Horton et al. (2011), Paolacci et al. (2010), Amir et al. (2012)). Bohannon (2016) claims that most of the participants on mTurk are professional experiment takers just as in many lab settings.

riety in highest completed education from high school to graduate school. The median participant is 33 years old and holds an undergraduate degree.

Participants' earnings were a \$0.50 show-up fee plus an average outcome-implied bonus of about \$1.06 for this 10−15 minutes task. Thus, they earned between \$6.22 ($= 1.56/10 \times 60$) and \$9.33 ($= 1.56/15 \times 60$) per hour on average.

Table 2.1 provides an overview of the Senders' descriptive statistics in both games.[17]

Table 2.1: Descriptive statistics

	Mean	St. dev.	Min	Med	Max	Observations
Gender	0.56	0.50	0.00	1.00	1.00	199
Age	34.56	11.23	18.00	32.00	76.00	199
Education	2.77	0.67	2.00	3.00	4.00	199
Proportion of lies	0.58	0.34	0.00	0.67	1.00	200
1st-order belief	0.51	0.25	0.00	0.50	1.00	200
Earnings (USD)	1.60	0.25	1.00	1.60	2.10	200
PV	4.04	0.97	1.78	3.89	7.00	199
Prosoc	3.62	1.56	1.00	3.50	7.00	199

This table shows descriptive statistics for Senders in the two games. Gender takes the value one for men and zero for women. Education denotes the highest completed education stage: 1: Elementary School, 2: High School, 3: Undergraduate School, 4: Graduate School. Proportion of lies refers to the ratio of incorrect to total messages transmitted by the Sender. The 1st-order belief corresponds to the Sender's assessment of the likelihood that the Receiver's action will match the message. Earnings include a fixed payment of \$0.50 and a variable compensation component. PV reflects an index of protected values for honesty. Prosoc is a variable capturing prosocial concerns. The full questionnaires underlying these variables are available in Appendix 2.B. One participant disconnected early resulting in demographic data for 199 individual Senders.

2.4.2 Overall Behavior

Overall, 58% of the 9000 messages sent are lies, meaning that the content of the message is contradictory to the true state of the world.[18] Senders in Game A tend to lie less than Senders in Game B, consistent with the original studies of Gneezy (2005) and Erat and Gneezy (2012), respectively. The mean percentage of lying decisions for a given Sender equals 53% in Game A and 64% in Game B, respectively. This extent of lying is generally above the fraction of lies in Gneezy (2005) who finds lying rates of 17% to 52% depending on the treatment, and Erat and Gneezy (2012) with a range from 43% to 76%, respectively. Both these studies use student participants and experiments that physically

[17] In this paper, we concentrate on Senders only and thus neglect Receiver behavior completely.

[18] We consider messages that contradict the true state of the world as lies regardless of the Sender's intention. As shown by Sutter (2009) a truthful message might also be used to deceive the other player if the transmitting player believes the Receiver not to follow her message.

take place in a laboratory. The share of following decisions per Receiver slightly varies across games and amounts to 90% in Game A and 85% in Game B on average. Figure 2.2 displays the distribution of Senders with respect to the proportion of lying decisions as well as the relationship between ECOST and the proportion of lies.

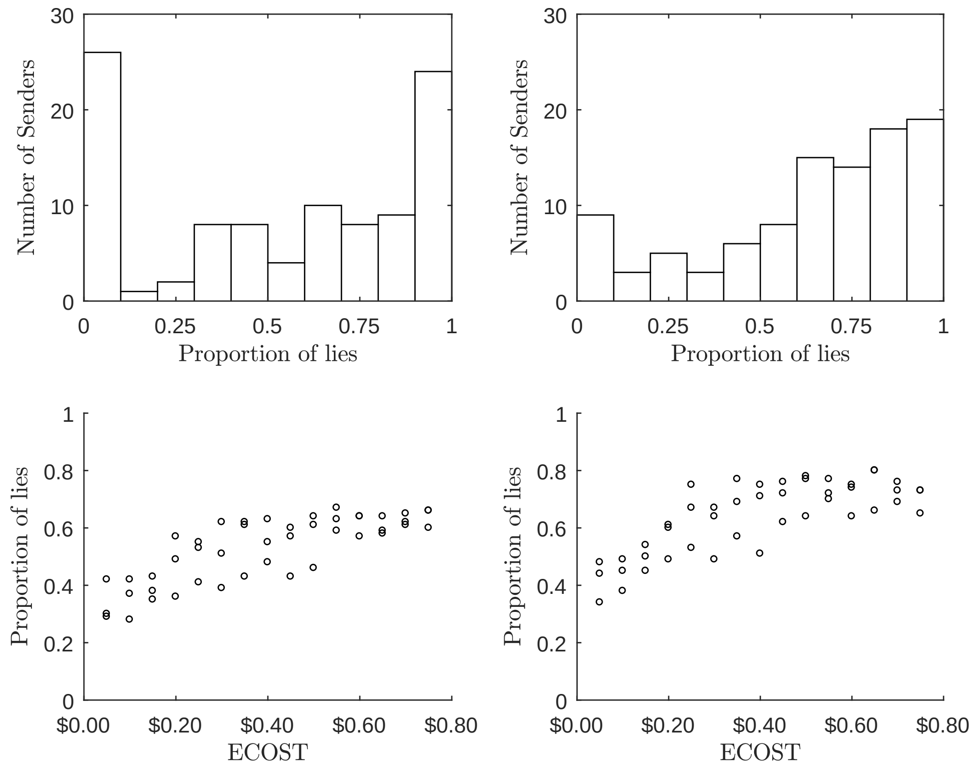


Figure 2.2: Descriptive lying behavior. The top two plots feature histograms of the proportion of lies. The bottom two plots illustrate the relationship of the proportion of lies with respect to ECOST. Each circle corresponds to one of the 45 different decision situations. The left plots correspond to Game A and the right plots to Game B, respectively.

Across games the histograms look slightly different. While the distribution in Game A is almost symmetric with peaks for very honest and dishonest behavior,[19] Game B exhibits right-skewness with a remarkable number of Senders frequently transmitting incorrect messages. Game B's histogram shows a clear step-wise ascent towards the highest frequency of telling lies. In contrast to Game A, a substantial number of Senders is lying in about 60% to 90% of all situations.

Each circle in the scatterplots in Figure 2.2 corresponds to an experimental round and reflects the ratio of lying Senders for a given decision situation. We note an increasing pattern with respect to ECOST. The higher the potential cost of transmitting a truthful

[19] The higher number of honest Senders in Game A compared to Game B could also indicate that some participants' preferences depend on the labels of the messages (in the spirit of Gneezy et al. (2016)). Identifying such types would require variation in framing within participants.

message the larger is the proportion of lying Senders (consistent with Gneezy (2005)). However, for higher ECOST the relationship flattens (consistent with findings in a non-strategic setting in Gibson et al. (2013)). Table 2.2 displays the relationship of lying and the decision environment as well as demographic variables.

Table 2.2: Determinants of the decision to lie

Dependent variable:	Game A		Game B		Combined	
Lie	(1)	(2)	(3)	(4)	(5)	(6)
ECOST	0.097***	0.246***	0.106***	0.133*	0.100***	0.186***
	(0.015)	(0.075)	(0.017)	(0.069)	(0.011)	(0.050)
Neg. Inequity	0.044***	0.018	0.041***	0.051**	0.042***	0.032**
	(0.008)	(0.019)	(0.009)	(0.021)	(0.006)	(0.014)
Belief R follows	0.401	0.398	-0.753	-0.752	-0.225	-0.223
	(0.785)	(0.788)	(0.575)	(0.576)	(0.493)	(0.493)
Gender	-0.099	-0.101	0.466*	0.466*	0.201	0.201
	(0.341)	(0.341)	(0.268)	(0.268)	(0.213)	(0.213)
Age	-0.014	-0.014	-0.008	-0.008	-0.009	-0.009
	(0.016)	(0.017)	(0.011)	(0.011)	(0.009)	(0.009)
Education	0.136	0.136	0.267	0.267	0.157	0.157
	(0.218)	(0.218)	(0.222)	(0.222)	(0.149)	(0.149)
PV	-0.959***	-0.670***	-0.528***	-0.475**	-0.690***	-0.524***
	(0.226)	(0.245)	(0.182)	(0.189)	(0.136)	(0.141)
ECOST $\times$ PV		-0.036**		-0.007		-0.021*
		(0.016)		(0.016)		(0.011)
Prosoc	-0.064	-0.122	-0.183*	-0.165	-0.121	-0.142*
	(0.119)	(0.110)	(0.108)	(0.105)	(0.079)	(0.075)
Neg. Inequity $\times$ Prosoc		0.007		-0.002		0.003
		(0.005)		(0.006)		(0.004)
Game					0.384	0.388
					(0.236)	(0.237)
Constant	3.113**	2.132	1.855	1.575	2.150**	1.544*
	(1.324)	(1.417)	(1.273)	(1.253)	(0.911)	(0.921)
Observations	4,455	4,455	4,500	4,500	8,955	8,955
Number of participants	99	99	100	100	199	199
Pseudo R^2	0.123	0.126	0.097	0.098	0.106	0.107

This table presents coefficients of logit regressions. The dependent variable is the binary variable Lie which takes the value one if the Sender tells a lie, and zero else. ECOST is the economic cost of stating the truth. Negative inequity refers to the difference $\Pi^R(\theta,\theta) - \Pi^S(\theta,\theta)$. Belief R follows denotes the Sender's first-order belief. Gender takes the value one for men and zero for women. Education takes the following values: 0: No schooling, 1: Elementary school, 2: High school, 3: Undergraduate degree, 4: Graduate school. PV reflects an index of protected values for honesty. Prosoc is a variable capturing prosocial concerns. The full questionnaires underlying these variables are available in Appendix 2.B. Game takes the value 0 for Game A and 1 for Game B, respectively. Columns (1) and (2) use data from Game A, columns (3) and (4) from Game B, and columns (5) and (6) from both games. Robust standard errors are obtained by clustering at the individual level and appear in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

Table 2.2 affirms the indicative evidence of Figure 2.2 and shows the proportion of lies to be increasing in ECOST. The increase in ECOST is diminishing for higher incentives. In addition, the difference in payoffs between Sender and Receiver influences lying. The less favorable the difference in payoffs for the Sender if reporting truthfully, i.e., the greater the negative inequity, the more likely she will lie. We would expect the coefficient on the Sender's first-order belief, i.e., the likelihood she assigns to the Receiver to match her message, to be significantly positive if it only entered the Sender's expected monetary payoff computation. The resulting coefficient indicates that the first-order belief does not only enter her expected monetary payoff calculation but also - at least for some - the moral costs of telling a lie. In addition, stronger protected values as well as prosocial concerns lead to more honesty. Finally, lying in Game B is significantly more pronounced than in Game A.

Our data confirm that the mTurk implementation of the Gneezy (2005) and Erat and Gneezy (2012) games with general population participants yields similar overall results as the laboratory setting with student participants. However, they also confirm that the two considered sender-receiver games, while similar in their fundamental structure lead to quite different overall behavior patterns across games.

2.5 Results

Based on the methodology explained in Section 2.3, we conduct the estimation, assign participants to types, and compute the in- and out-of-sample fits for all worlds under consideration. Appendix Table 2.D.2 and Table 2.D.4 provide a complete overview of all worlds and how well they fit the data in- as well as out-of-sample in Game A and B, respectively.[20] For our analysis we focus on the share of correctly fitted decisions. Section 2.5.1 considers model selection when choosing one of the five available types. Section 2.5.2 allows for multiple types. Section 2.5.3 studies the shares of types in the population and the demographic and other correlates of the probability with which a participant is assigned to being a particular type.

2.5.1 Selecting One Preference Type

We begin with model selection that involves selecting the best among the five available types. Thus, this special case only considers the five one-type worlds. Figure 2.3 shows the in- and out-of-sample fit for all these one-type worlds. Similarly, Table 2.3 and Table 2.4 display the computed in- and out-of-sample performances in Game A and B, respectively. In addition, the accuracies are tested for statistical significance.

Figure 2.3 highlights that among all five one-type worlds, one world features substantially worse in-sample and out-of-sample accuracy, namely the EC world. The standard

[20] The standard errors of the fits for all worlds are displayed in Appendix Table 2.D.3 and Table 2.D.5 for Games A and B, respectively.

economic model does really poorly at organizing data as participants seem not to follow the strategy to always lie.[21] The bad in- and out-of-sample fit of EC is unsurprising given that we only estimate the parameter of erroneous play λ_i and that EC is nested in all other types.

Our main interest is in comparing the four worlds that have at least one additional parameter. These worlds explain behavior better. Moreover, in-sample fits are always better than the respective out-of-sample fits. However, lying costs (LC) and social norms (SN) perform better than both inequity aversion (IA) and reference-dependent preferences (RD). LC explains behavior significantly best in-sample when approximately less than half of the full sample is used to estimate and classify participants, and SN performs best in-sample when more than half of the sample is used. However, an approach based on reputation and social norms (SN) does fit the data best out-of-sample, regardless of the number of decisions in the in-sample part. There are no substantial differences between Game A and B.

This first analysis highlights why many papers that analyze lying in the context of strategic information transmission suggest lying cost, reputation, or norm-based explanations to explain their findings as these are the types that do really well in-sample. Based on this baseline analysis with five one-type worlds we are already able to observe some of our key insights: First, the quality of the in-sample fit is decreasing in the number of decisions used to estimate parameters and to assign participants to types. The more decisions, the tougher it is to explain each and every choice. Second, the quality of the out-of-sample fit is usually increasing in the number of decisions used to estimate parameters and to assign participants to types. More decisions make it easier to estimate and classify participants into types implying a better classification and hence better out-of-sample predictions. Third, worlds that fit the data well in-sample do not necessarily perform well out-of-sample and vice versa. Finally, the size of the in-sample part of the data is important for determining the best possible worlds.

[21] If one allowed for all possible first-order beliefs instead, the Sender's strategy under EC would be to lie whenever the expected monetary payoff of lying is greater than of telling the truth, i.e., whenever the participant's first-order belief is greater than one half in Game A or greater than one sixth in Game B, respectively. Results for the case of self-reported beliefs are available on request.

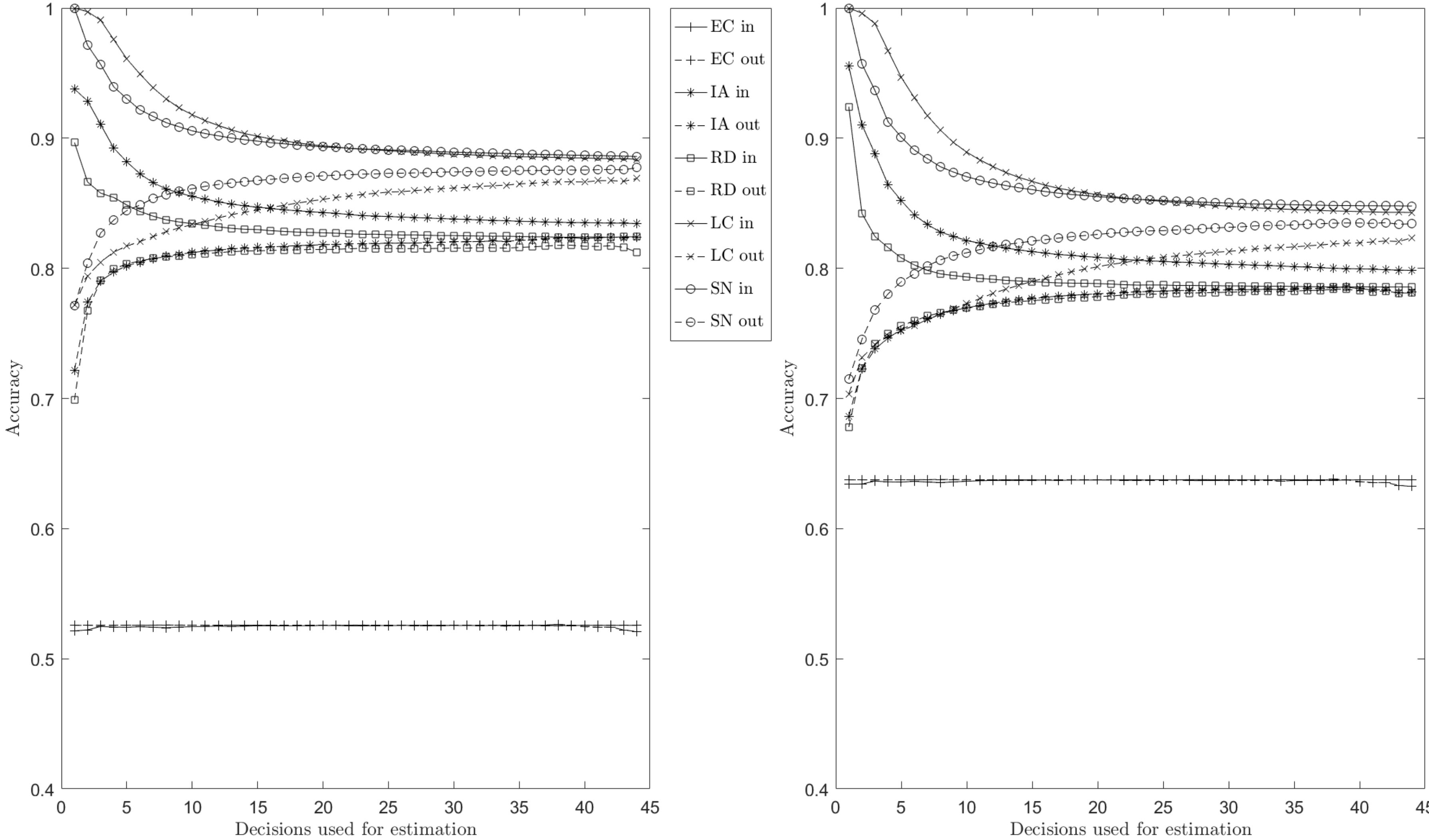


Figure 2.3: In- and out-of-sample performance of one-type worlds. Solid lines reflect in-sample accuracy, dashed lines out-of-sample accuracy. EC denotes the standard economic model, IA inequity aversion, RD reference-dependent preferences, LC lying costs, and SN reputation or social norms. The left plot correspond to Game A, and the right plot to Game B, respectively. The accuracies are derived based on $M = 1000$ random draws.

Table 2.3: Performance of one-type worlds in Game A

	In-sample accuracy Decisions in-sample				
	1	11	22	33	44
EC	0.5214	0.5249	0.5253	0.5256	0.5257
IA	0.9377***	0.8531***	0.8413***	0.8368***	0.8344***
RD	0.8966***	0.8330***	0.8265***	0.8246***	0.8242***
LC	1.0000***	0.9134***	0.8921***	0.8864***	0.8838***
SN	1.0000***	0.9035***	0.8923	0.8885***	0.8859***
	Out-of-sample accuracy Decisions in-sample				
	1	11	22	33	44
EC	0.5257	0.5258	0.5258	0.5256	0.5208
IA	0.7216***	0.8133***	0.8186***	0.8209***	0.8240***
RD	0.6989***	0.8114***	0.8150***	0.8158***	0.8123***
LC	0.7717	0.8360***	0.8555***	0.8634***	0.8690***
SN	0.7711***	0.8630***	0.8719***	0.8748***	0.8774***

This table presents the computed in-sample and out-of-sample accuracies in Game A. Decisions in-sample corresponds to the number of decisions used to estimate the one-type worlds. For each accuracy we run a two-tailed t-test against the next lower accuracy for a given number of decisions in-sample. We test the out-of-sample accuracies similarly. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively. Accuracies are depicted in Table 2.D.2 and the corresponding standard errors in Table 2.D.3.

Table 2.4: Performance of one-type worlds in Game B

	In-sample accuracy Decisions in-sample				
	1	11	22	33	44
EC	0.6342	0.6368	0.6372	0.6374	0.6374***
IA	0.9553***	0.8193***	0.8068***	0.8021***	0.7984***
RD	0.9242***	0.7923***	0.7874***	0.7862***	0.7856***
LC	1.0000***	0.8833***	0.8540	0.8460***	0.8429***
SN	1.0000***	0.8677***	0.8536***	0.8493***	0.8479***
	Out-of-sample accuracy Decisions in-sample				
	1	11	22	33	44
EC	0.6374	0.6375	0.6375	0.6371	0.6323
IA	0.6863***	0.7715	0.7814***	0.7840	0.7819
RD	0.6781***	0.7709***	0.7794***	0.7824***	0.7811***
LC	0.7035***	0.7767***	0.8043***	0.8156***	0.8234***
SN	0.7147***	0.8144***	0.8276***	0.8328***	0.8342***

This table presents the computed in-sample and out-of-sample accuracies in Game B. Decisions in-sample corresponds to the number of decisions used to estimate the one-type worlds. For each accuracy we run a two-tailed t-test against the next lower accuracy for a given number of decisions in-sample. We test the out-of-sample accuracies similarly. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively. Accuracies are depicted in Table 2.D.4 and the corresponding standard errors in Table 2.D.5.

2.5.2 Considering Type Combinations

We now allow for richer worlds, not forcing everybody to be of the same type. Because the number of different worlds is large we restrict our attention to unique and well performing worlds. We proceed as follows: First, we eliminate all worlds that are equivalent to a world which allows for fewer types. Then, we compare the remaining worlds and begin to eliminate dominated worlds.

Worlds are considered equivalent if all participants in two worlds (with a different number of available types) are assigned to the same types. This implies that at least one type is redundant and contains no members. As a result the world in which a given type is redundant can be eliminated in favor of a smaller world that considers the same types except the redundant.

A world is considered dominated if its accuracy is strictly smaller than another world's for all number of decisions used to estimate and classify participants greater than 3. We separately analyze dominance in the in-sample and the out-of-sample dimension.[22] Worlds that are dominated, in-sample or out-of-sample, are eliminated. As we refine separately for in-sample and out-of-sample fits, it is very likely that some worlds will be dominated in one of the two dimensions only.

After this procedure, we are left with 4 in-sample and 2 out-of-sample worlds in Game A (using $M = 1000$ different random selections for the specific decisions used to estimate parameters and assign participants to types), and 3 in-sample and 3 out-of-sample worlds in Game B ($M = 1000$), respectively. Figure 2.4 provides the in- and out-of-sample performance of all worlds that are not dominated in at least one dimension for both games. In addition, the fit of the world that leads to the highest combined accuracy, i.e., in-sample plus out-of-sample, is drawn.

Overall, the figures for the two experiments are quite similar. If the number of decisions used to estimate and classify people into types is comparably small (Game A: < 11, Game B: < 14) a simple one-type world featuring only participants with constant, linear or convex lying costs explains behavior significantly best in-sample. Once the size of the in-sample part of the data increases a two-type world with lying costs and reputation (or a three-type world that also includes the reference-dependent type) outperforms. More elaborated worlds with four- or five types are not able to further improve the in-sample accuracy.

[22] The in-sample fit of most models, if only few decision are used to estimate the parameters, is very similar and only differs due to numerical reasons. In order to control for these numerical differences, we neglect the fits derived from only 1, 2, or 3 decisions for our elimination, i.e., we only eliminate worlds that are dominated if 4 to 44 decisions were used to estimate and classify participants into types.

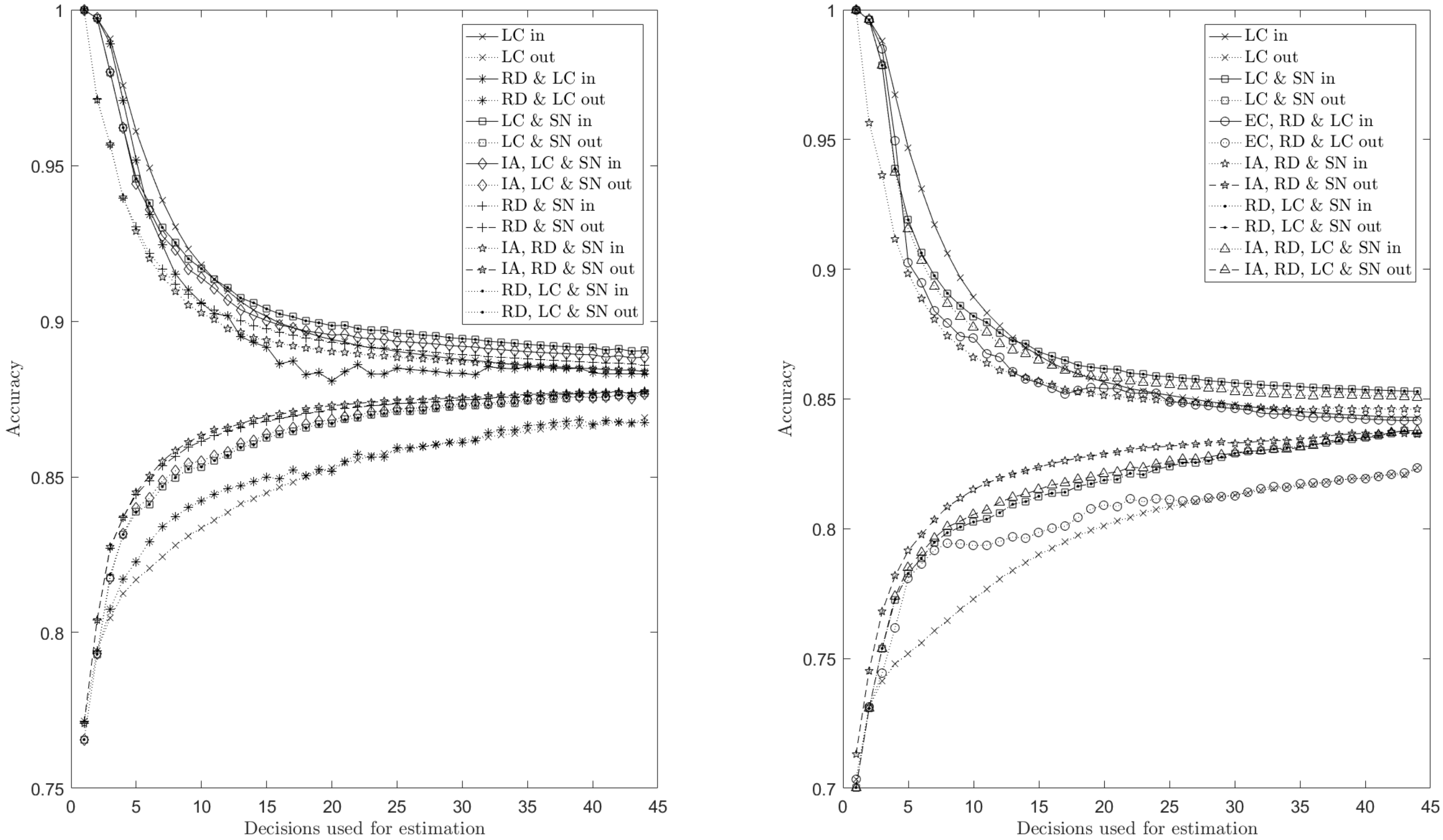


Figure 2.4: In- and out-of-sample performance of non-dominated worlds. In Game A (left-hand side) the following worlds are not dominated in-sample: LC; RD & LC; LC & SN; IA, LC & SN; and out-of-sample: RD & SN; IA, RD & SN. In Game B (right-hand side) the following worlds are not dominated in-sample: LC; LC & SN; EC, RD & LC; and out-of-sample: IA, RD & SN; RD, LC & SN; IA, RD, LC & SN. World RD, LC & SN explains behavior best when summing in- and out-of-sample performance. EC reflects the standard economic model, IA inequity aversion, RD reference-dependent preferences, LC lying costs, and SN reputation or social norms. Worlds which are not dominated in-sample feature solid lines and the corresponding out-of-sample fits dotted lines. Worlds which are not dominated out-of-sample feature dashed lines and the corresponding in-sample fits dotted lines. The accuracies are derived based on $M = 1000$ random selections.

Out-of-sample the story is different. The simple LC only world does really poorly in fitting the behavioral data out-of-sample compared to the other worlds regardless of the sample size used for the estimation. This finding raises a caveat for behavioral research and specifically the experimental literature on lying. As experimental studies in the corresponding literature mostly base their results on only few decisions per participant, it is very likely that they find a lying costs model to organize the data well. However, as shown in Figure 2.4, not only will a model considering lying costs no longer do well in-sample once more data is collected, but also - and much more importantly - it will do a terrible job compared to other models in predicting future lying behavior.

We find a world with types IA, RD, and SN to explain behavior best out-of-sample most of the time. It reaches an out-of-sample accuracy of 0.85 in Game A and 0.80 in Game B already after 6 decisions in the estimation part of the data. While this world does significantly outperform the best-performing in-sample worlds, it would never be chosen if a study only accounted for the accuracy of the in-sample fit. Adding in-sample and out-of-sample performance, we find a world with RD, LC, and SN to explain behavior best in both games.

Figure 2.4 reveals a dilemma of model selection. The quality of the in-sample fit does not say anything about how well the same model will be able to explain future behavior. Similarly, worlds that perform comparably poorly in-sample might very well do a pretty good job in organizing future data sets. As a consequence, it is crucial that researchers are very cautious when deciding about a preferred model based on an in-sample methodology only as it might provoke misleading modeling choices.

2.5.3 World Characteristics

This subsection analyzes the shares of the population assigned to each type and the characteristics that explain the probability of a participant being assigned to a given type. For this analysis, we use the five-type model as a benchmark in order to have one that can be analyzed with respect to all types and in both games even though some worlds that feature less types can explain behavior slightly better. In Figure 2.5, we investigate the classification of participants. We plot the distribution of types, i.e., the share of each type within the population in both games.

A desirable property of model selection is for the distribution of types in the population to be similar no matter what the exact decision environment is. Finding stable shares in different experimental data sets with different participants and somewhat different games suggests that the proposed world does a reasonable job in explaining how people in the population behave. Figure 2.5 provides indicative evidence that the share of each type within the whole population is quite stable across different games.

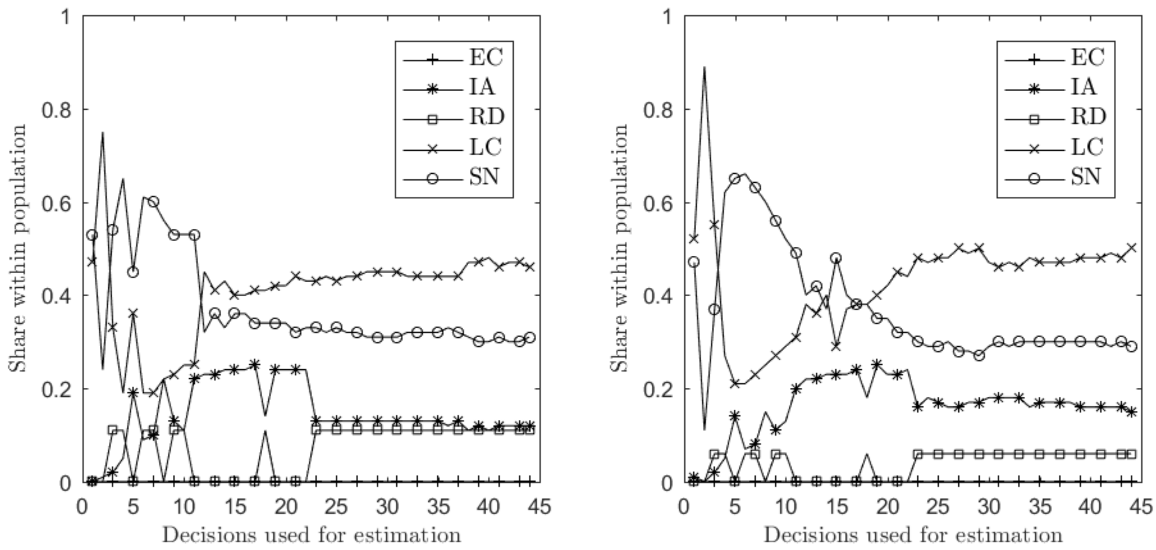


Figure 2.5: Distribution of types within the population with respect to the number of decisions used to estimate and classify participants into types in the five-type world. EC reflects the standard economic model, IA inequity aversion, RD reference-dependent preferences, LC lying costs, and SN reputation or social norms. The left plot correspond to Game A, and the right plot to Game B, respectively. The distributions are derived based on $M = 1000$ random draws.

We find LC and SN to make up the largest shares in the population. This is in line with Figure 2.3 which showed that these two types are also the ones that can organize data quite well in a one-type world. Moreover, Figure 2.5 highlights that the distribution of types also depends on the size of the sample data used to estimate and classify participants into types.[23] Again, using only a small number of decisions might imply very misleading conclusions.[24]

Next, we relate the probability of belonging to a specific type to demographic characteristics such as gender, age, and education, and protected values and prosocial concerns. The probability of belonging to a given type is computed as the average of a binary assignment variable which takes the value one if a participant is assigned to the given type in a specific draw and for a given number of decisions, and zero else. We derive the averages of these binary assignment variables over all 44 possible number of decisions and the $M = 1000$ selections.

Table 2.5 displays the results of the corresponding regressions for the five-type world.[25]

[23] The one exception is type EC, to which no participant is assigned.

[24] Keep in mind that we consider binary choices only. If one considered settings with a richer message space instead, collecting information about fewer decisions might be sufficient.

[25] Comparing the results for the benchmark world with the best in-sample, out-of-sample, or combined world, i.e., a world consisting of types LC and SN in-sample, a world with types IA, RD, and SN out-of-sample, or RD, LC & SN which features the best combined accuracy, yields similar results indicating that people are broadly assigned to similar categories.

Table 2.5: Type classification and personal characteristics

Dependent variable:	Type EC (1)	Type IA (2)	Type RD (3)	Type LC (4)	Type SN (5)
Belief R follows		-0.145	0.122*	0.070	-0.046
		(0.090)	(0.066)	(0.113)	(0.102)
Gender		0.024	-0.025	-0.016	0.016
		(0.040)	(0.028)	(0.055)	(0.054)
Age		-0.001	-0.001	0.002	0.001
		(0.001)	(0.001)	(0.002)	(0.002)
Education		0.051*	-0.008	-0.027	-0.016
		(0.029)	(0.020)	(0.041)	(0.037)
PV		-0.035	-0.037**	0.066**	0.005
		(0.023)	(0.016)	(0.031)	(0.029)
Prosoc		-0.006	-0.022***	0.013	0.016
		(0.016)	(0.008)	(0.019)	(0.018)
Game		-0.009	-0.020	0.034	-0.006
		(0.041)	(0.032)	(0.059)	(0.056)
Constant		0.269	0.319***	0.080	0.331
		(0.175)	(0.118)	(0.231)	(0.232)
Observations	199	199	199	199	199
R-squared		0.048	0.070	0.034	0.007

This table presents coefficients of OLS regressions. The dependent variables are frequencies a given participant is assigned to the respective type. The frequency a participant is assigned to a given type corresponds to the average over all number of decisions and random selections. Column (1) refers to the standard economic model (however, no participant is assigned to EC), column (2) to social preferences, column (3) to reference-dependent preferences, column (4) to lying costs, and column (5) to reputation and social norms. Belief R follows indicates the Sender's first-order belief with respect to the likelihood that the Receiver chooses what the Sender has sent in her message. Gender takes the value one for men and zero for women. Education takes the following values: 0: No schooling, 1: Elementary school, 2: High school, 3: Undergraduate degree, 4: Graduate school. The measurement of PV as well as Prosoc is described in Appendix 2.B. Game takes the value one in Game B, and zero else. Robust standard errors appear in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively.

The empty first column in the table indicates that no participants are assigned to EC regardless of the number of decisions used to estimate and classify participants into types. Columns (2) to (5) show that demographic characteristics hardly have any influence on the type to which a participant is assigned. One exception is that well-educated participants are more likely assigned to IA. Participants with strong protected values and prosocial concerns are significantly less often classified as RD, and high-PV individuals are more often classified as LC. Finally, we find no impact of the specific game on the classification, highlighting that our classification is robust to changes in the strategic decision environment.

2.6 Conclusion

We collect and make available a comprehensive data set of lying decisions in strategic information transmission settings. Using these data, we deliver two substantive results and one methodological accomplishment: First, we classify people into five different types according to their behavior. We analyze how models that allow for different types fit experimental data in-sample as well as out-of-sample. Second, our paper reveals the quality of the in-sample fit does not say anything about how well the same model will be able to explain behavioral data out-of-sample. Similarly, worlds that perform poorly in-sample might very well do a pretty good job in organizing future data sets. As a consequence, it is crucial that researchers are very cautious when deciding about a preferred model based on an in-sample methodology only as it might provoke misleading modeling choices especially when using a small number of decisions. Third, in terms of methods, the paper provides a framework that can be employed in other instances where researchers are interested in assessing plausible preference specifications.

Since we run our analysis in two different strategic information transmission environments we are confident that our key findings are robust to other applications of similar settings. Further research could analyze parameter stability and how the number of rounds used for the estimation influences the persistence of the parameter estimates within types. We find indicative evidence that the distribution of the individual parameter estimates within types and the share of each type within the population are quite stable across different strategic decision situations.

Appendices

2.A Appendix A: Payoff Matrices

Table 2.A.1: Payoffs

Ξ	(Π^S, Π^R)		Ξ	(Π^S, Π^R)		Ξ	(Π^S, Π^R)	
	$a = \theta$	$a \neq \theta$		$a = \theta$	$a \neq \theta$		$a = \theta$	$a \neq \theta$
	$(24, 25)$	$(25, 10)$		$(21, 22)$	$(27, 12)$		$(19, 20)$	$(30, 15)$
\$0.05	$(20, 28)$	$(21, 13)$	\$0.30	$(18, 26)$	$(23, 15)$	\$0.55	$(16, 24)$	$(27, 19)$
	$(17, 32)$	$(18, 17)$		$(14, 29)$	$(20, 19)$		$(12, 27)$	$(23, 22)$
	$(23, 24)$	$(25, 10)$		$(21, 22)$	$(28, 13)$		$(19, 20)$	$(31, 16)$
\$0.10	$(20, 28)$	$(22, 14)$	\$0.35	$(18, 26)$	$(25, 17)$	\$0.60	$(15, 23)$	$(27, 19)$
	$(16, 31)$	$(18, 17)$		$(14, 29)$	$(21, 20)$		$(12, 27)$	$(24, 23)$
	$(23, 24)$	$(26, 11)$		$(21, 22)$	$(29, 14)$		$(18, 19)$	$(31, 16)$
\$0.15	$(20, 28)$	$(23, 15)$	\$0.40	$(17, 25)$	$(25, 17)$	\$0.65	$(14, 22)$	$(27, 19)$
	$(16, 31)$	$(19, 18)$		$(14, 29)$	$(22, 21)$		$(11, 26)$	$(24, 23)$
	$(23, 24)$	$(27, 12)$		$(20, 21)$	$(29, 14)$		$(17, 18)$	$(31, 16)$
\$0.20	$(19, 27)$	$(23, 15)$	\$0.45	$(16, 24)$	$(25, 17)$	\$0.70	$(14, 22)$	$(28, 20)$
	$(16, 31)$	$(20, 19)$		$(13, 28)$	$(22, 21)$		$(10, 25)$	$(24, 23)$
	$(22, 23)$	$(27, 12)$		$(19, 20)$	$(29, 14)$		$(17, 18)$	$(32, 17)$
\$0.25	$(18, 26)$	$(23, 15)$	\$0.50	$(16, 24)$	$(26, 18)$	\$0.75	$(13, 21)$	$(28, 20)$
	$(15, 30)$	$(20, 19)$		$(12, 27)$	$(22, 21)$		$(10, 25)$	$(25, 24)$

This table summarizes all 45 payoff matrices in experimental units. One point in the experiment corresponds to \$0.05 real money. Ξ refers to the implied ECOST in \$ and is not shown to participants. The tuple (Π^S, Π^R) denotes the Sender's and the Receiver's payoff, respectively. The two possible outcomes are $a = \theta$ (i.e., action and state coincide) and $a \neq \theta$ (i.e., action and state differ) and depend on the Receiver's choice a only.

2.B Appendix B: Experimental Instructions

In general, the wording is kept very similar to the underlying studies of Gneezy (2005) and Erat and Gneezy (2012).

2.B.1 Game A

2.B.1.1 Sender

After a participant has accepted the HIT on mTurk and is assigned to be a Sender, we provide detailed instructions about Game A:

Instructions

In this study you will play 45 *rounds of a decision making experiment. Please read the instructions carefully. You may earn a considerable sum of money, depending on the decisions you make in the experiment.*

In the end, one of the 45 *rounds will be randomly selected and you will be randomly matched with another (anonymous) participant. Neither of you will ever know the identity of the other. The money you earn in this particular round will be paid to you at the end of the experiment in addition to a fixed amount of* \$0.50 *for participation. The conversion rate is set such that* 20 *points in the experiment correspond to* \$1.00 *(i.e.,* \$0.05 *per point in the experiment).*

In every round, two possible monetary payments are available to you and your counterpart in the experiment. For example, the two payment options might be:

- *Option X:* 35 *points to you and* 25 *points to the other player*
- *Option Y:* 28 *points to you and* 32 *points to the other player*

Note that the monetary payments of the two options will change from round to round.

The choice rests with the other player who will have to choose either Option X or Option Y. The only information your counterpart will have is information sent by you in a message. That is, he or she will not know the monetary payments associated with each option.

We will ask you to choose one of the following two possible messages, which you will send to your counterpart:

- *Message 1: "Option X will earn you more money than Option Y."*
- *Message 2: "Option Y will earn you more money than Option X."*

We will show the other participant your message, and ask him or her to choose either X or Y. To repeat, your counterpart's choice will determine the payments in the experiment.

However, your counterpart will never know what sums were actually offered in the option not chosen (that is, he or she will not be able to infer whether your message was true or not). Moreover, he or she will never know the sums to be paid to you according to the different options.

We will pay the two of you according to the choice made by your counterpart.

For your convenience, these instructions will remain available to you on all subsequent screens of this study.

The instructions are followed by two understanding questions which check the Sender's knowledge of the game structure and the payoff derivation. The Sender is informed about the correctness of her answers and, if the answer is incorrect, shown the correct solution with an explanation.

Then, the first round starts:

Round 1 of 45

We ask you to send a message to the other player. This message is the only information your counterpart will receive. He or she will not be informed about the monetary consequences of each option. In this round the monetary consequences are as follows:

- *If your counterpart chooses Option X, you will earn* 19 *points and your counterpart will earn* 20 *points.*
- *If your counterpart chooses Option Y, you will earn* 29 *points and your counterpart will earn* 14 *points.*

I choose to send:

- *Message 1: "Option X will earn you more money than Option Y."*
- *Message 2: "Option Y will earn you more money than Option X."*

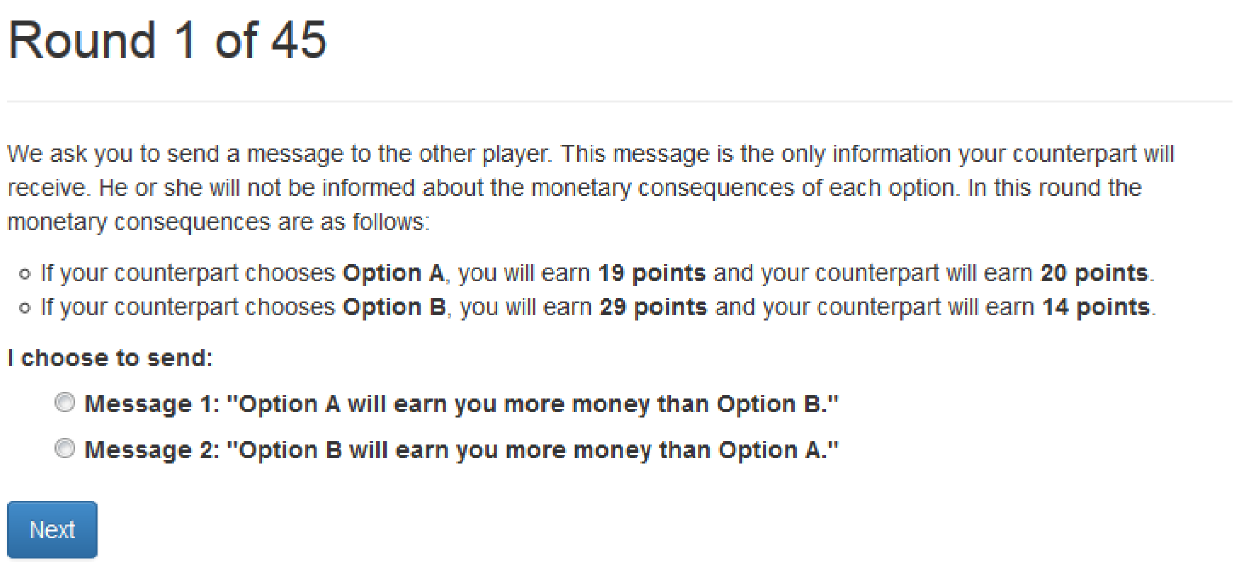
The exact screen the Sender faces is depicted in Figure 2.B.1.

The subsequent rounds are formatted similarly but feature different payoff matrices.

In a next step, we elicit the Sender's first-order belief by asking: *Out of* 100 *possible counterparts, how many do you think will follow your messages (i.e., choose the option you have recommended)?*

In the questionnaire we ask demographics (i.e., gender, age, education) and use Tanner et al. (2009)'s protected values survey. We adjusted the questions to our context:

1. *Consider the choices of the player who sent a message. To what extent do you judge the decision to send something other than the truthful message as*
    - *1: Very immoral ... 7: Very moral*



Round 1 of 45

We ask you to send a message to the other player. This message is the only information your counterpart will receive. He or she will not be informed about the monetary consequences of each option. In this round the monetary consequences are as follows:

- If your counterpart chooses **Option A**, you will earn **19 points** and your counterpart will earn **20 points**.
- If your counterpart chooses **Option B**, you will earn **29 points** and your counterpart will earn **14 points**.

I choose to send:

Message 1: "Option A will earn you more money than Option B."

Message 2: "Option B will earn you more money than Option A."

Next

Figure 2.B.1: Decision situation for a Sender in Game A

- *1: Not at all praiseworthy … 7: Very praiseworthy*
- *1: Not at all blameworthy … 7: Very blameworthy*
- *1: Not at all outrageous … 7: Very outrageous*
- *1: Not at all acceptable … 7: Very acceptable*

2. *Individuals have an opportunity to choose which information they provide to their audience, and they can modify information they obtain before passing this information on to others. Some view such modification as a violation of truthfulness; others regard it as acceptable protection of personal interests. What do you think about the value of truthfulness in such a situation? Truthfulness is about something…*

   - *… that one should not sacrifice, no matter what the (material or other) benefits. [1: Strongly Disagree … 7: Strongly Agree]*
   - *… for which I think it is right to make a cost-benefit analysis. [1: Strongly Disagree … 7: Strongly Agree]*
   - *… that cannot be measured in monetary terms. [1: Strongly Disagree … 7:*
   - *… about which I can be flexible if the situation demands it. [1: Strongly Disagree … 7: Strongly Agree]*

In addition, we elicit players' prosocial concerns on a 7-point Likert scale. While the second (Manipulative) and the fourth (Hurting) question make up the prosocial concerns measure, questions one and three serve as control questions.

Suppose that the player who must choose either Option X or B earns more from choosing Option Y. To what extent do you judge the decision to send Message 1: "Option X will earn you more money than Option Y." as:

- *1: Very dishonest ... 7: Very honest*
- *1: Very manipulative ... 7: Not at all manipulative*
- *1: Strongly associated with personal gains ... 7: Not at all associated with personal gains*
- *1: Strongly hurting the other participant ... 7: Not at all hurting the other participant*

Finally, Figure 2.B.2 shows the final screen of the experiment before being returned to mTurk.[26]

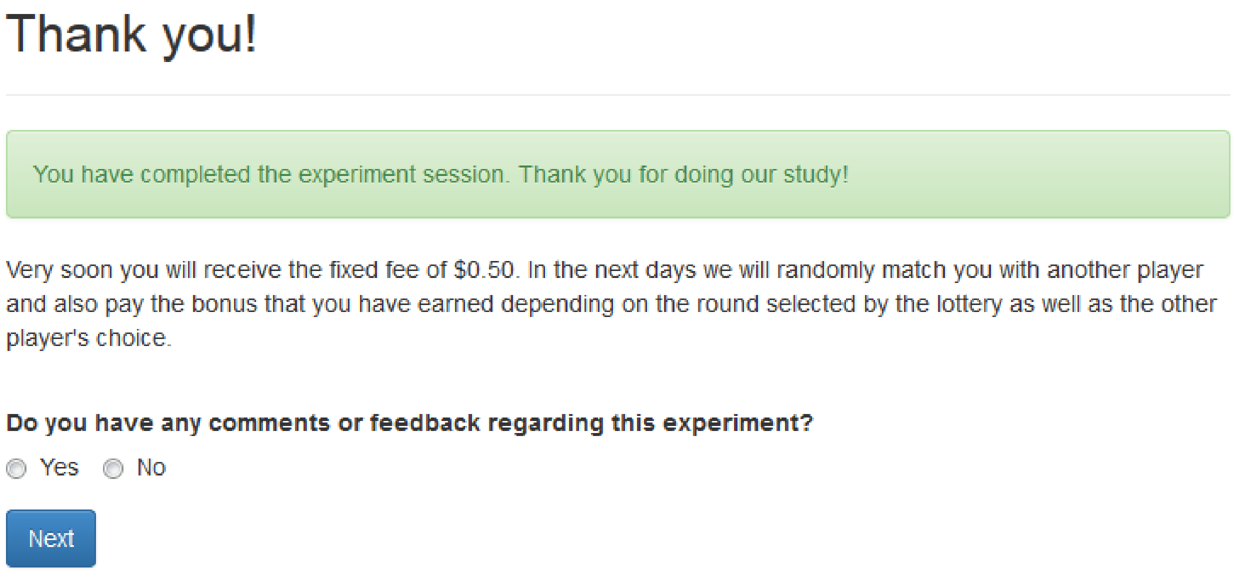


Figure 2.B.2: End of the experiment

2.B.1.2 Receiver

A participant who is assigned to be a Receiver in Game A is provided with the following instructions:

[26] In fact, for a first pilot experiment, we have directly matched participants and shown them the variable payoffs they achieved. However, this caused, on the one hand, unsatisfactory waiting periods for some participants due to a delayed start of others, and, on the other hand, administrative effort for the experimenter due to players disconnecting in between.

Instructions

This is a short experiment in decision making. Please read the instructions carefully. You may earn a considerable sum of money, depending on the decisions you make in the experiment.

In this experiment you will be randomly matched with another (anonymous) participant. Neither of you will ever know the identity of the other. The money you earn in this game will be paid to you in addition to a fixed amount of $0.50 for participation.

Two possible monetary payments are available to you and your counterpart in the experiment. The payments depend upon the option chosen by you. We showed the two payment options to your counterpart. The only information you will have is the message your counterpart sends to you.

Two possible messages could be sent:

- *Message 1: "Option X will earn you more money than Option Y."*
- *Message 2: "Option Y will earn you more money than Option X."*

Upon observing what message your counterpart sent, you may choose either Option X or Option Y. Your choice will determine the payments in the experiment. You will never know what sums were actually offered in the option not chosen (that is, if the message sent by your counterpart was true or not). Moreover, you will never know the sums your counterpart could be paid with the other option.

We will pay the two of you according to the choice you make.

For your convenience, these instructions will remain available to you on all subsequent screens of this study.

Right after the instructions, the Receiver's decision screen is shown asking for the Receiver's strategies conditional on all possible messages (see Figure 2.B.3):

Decisions

Your counterpart will send you a message. For every possible message we ask you to choose one of the two options. Your choice will determine the payments to you and your counterpart.

Decision 1: Suppose your counterpart sent you: Message 1: "Option X earns you more money than Option Y."

Given this message I would choose:

- *Option X*
- *Option Y*

Decision 2: Suppose your counterpart sent you: Message 2: "Option Y earns you more money than Option X."

Given this message I would choose:

- *Option X*
- *Option Y*

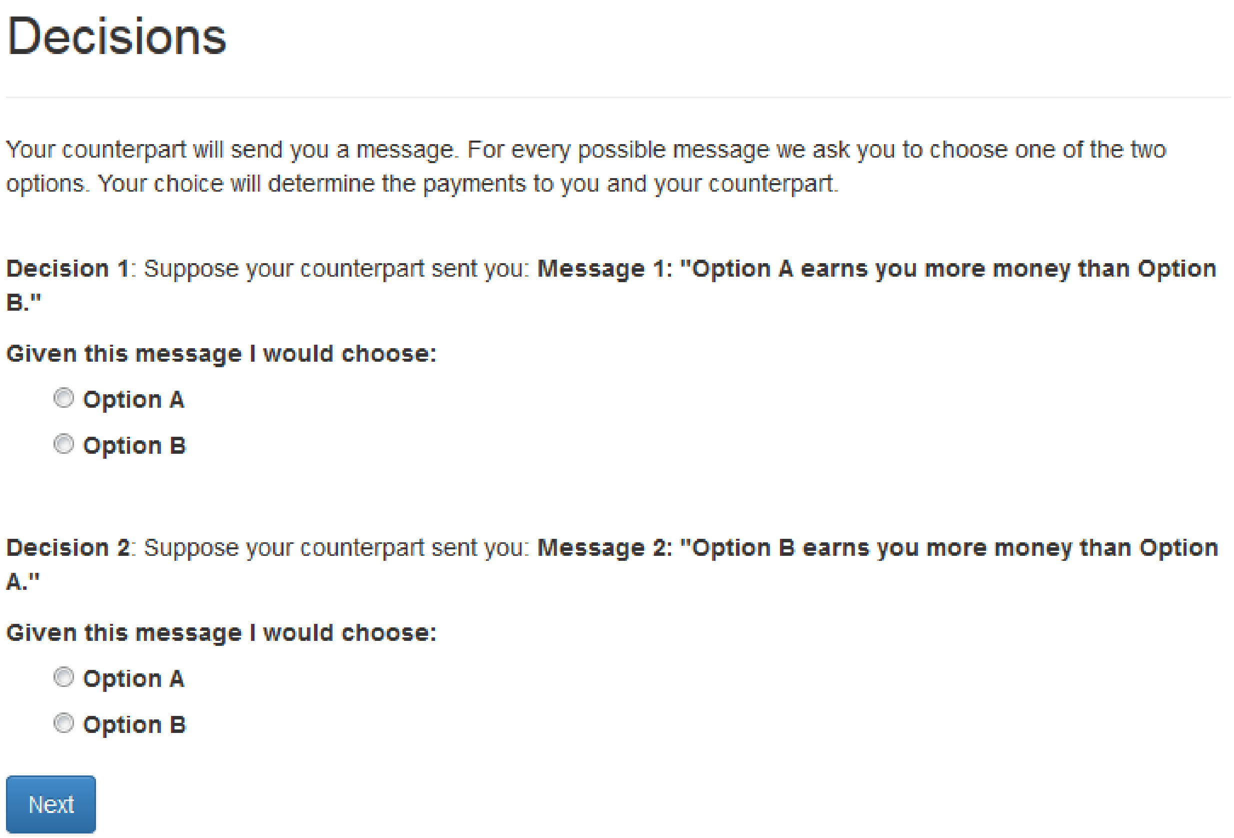


Figure 2.B.3: Decision situation for a Receiver in Game A

In a next step, we elicit the Receiver's first-order belief by asking: *Out of* 100 *possible counterparts, how many do you think will send you a truthful message (i.e., report the option which will earn you more)?*

Finally, a questionnaire and a final screen similar to the Sender's close the experiment.

2.B.2 Game B

2.B.2.1 Sender

A participant who is assigned to be a Sender is provided with detailed instructions about Game B:

Instructions

In this study you will play 45 *rounds of a short decision making experiment. Please read the instructions carefully. You may earn a considerable sum of money, depending on the decisions you make in the experiment.*

In the end, one of the 45 *rounds will be randomly selected and you will be randomly matched with another (anonymous) participant. Neither of you will know the identity of the other. The money you earn in this particular round will be paid to you at the end of the experiment in addition to a fixed amount of* \$0.50 *for participation. The conversion rate is set such that* 20 *points in the experiment correspond to* \$1.00 *(i.e.,* \$0.05 *per point in the experiment).*

Before every round, we will roll a 6-sided die. The other participant will not be informed about the outcome of the die roll. However, he or she will be told that you have been informed about the outcome of the die roll. In every round, you will be asked to send a message to the other participant. The message will correspond to a number from 1 to 6. There are six possible messages:

- *Message 1: "The outcome from the roll of the 6-sided die is 1."*
- *Message 2: "The outcome from the roll of the 6-sided die is 2."*
- *Message 3: "The outcome from the roll of the 6-sided die is 3."*
- *Message 4: "The outcome from the roll of the 6-sided die is 4."*
- *Message 5: "The outcome from the roll of the 6-sided die is 5."*
- *Message 6: "The outcome from the roll of the 6-sided die is 6."*

Your message will be shown to the other participant, and then he or she will be asked to choose a number between 1 and 6.

The choice of the number by the other participant will determine the payments in the experiment. For example:

- *If he or she chooses the actual outcome of the roll of the dice, then you will receive* 35 *points and he or she will receive* 25 *points.*
- *If he or she chooses a number different than the actual outcome, you will receive* 28 *points and he or she will receive* 32 *points.*

Note that the monetary payments will change from round to round.

Finally, only you will be informed of the particular monetary value connected to each message. The other participant will not be informed of these monetary values. However, he or she will be told that you have been informed of the monetary value connected to each message.

For your convenience, these instructions will remain available to you on all subsequent screens of this study.

The instructions are followed by two understanding questions which check the Sender's knowledge of the game structure. The Sender is informed about the correctness of her answers and, if the answer is incorrect, shown the correct solution with an explanation.

Then, the first round starts:

Round 1 of 45

Before starting this round, we have rolled a 6-sided die, and obtained the outcome 2. We ask you to send a message to the other participant. This message is the only information he or she will receive. He or she will not be informed about the monetary consequences of each option. In this round the monetary consequences are as follows:

- *If he or she chooses the actual outcome of the roll of the dice, then you will receive 19 points and he or she will receive 20 points.*

- *If he or she chooses a number different than the actual outcome of the roll of the dice, you will receive 29 points and he or she will receive 14 points.*

The message I will send is:

- *Message 1: "The outcome from the roll of the 6-sided die is 1."*

- *Message 2: "The outcome from the roll of the 6-sided die is 2."*

- *Message 3: "The outcome from the roll of the 6-sided die is 3."*

- *Message 4: "The outcome from the roll of the 6-sided die is 4."*

- *Message 5: "The outcome from the roll of the 6-sided die is 5."*

- *Message 6: "The outcome from the roll of the 6-sided die is 6."*

The exact screen the Sender faces is depicted in Figure 2.B.4.

The subsequent rounds are formatted similarly but feature different payoff matrices.

In a next step, we elicit the Sender's first-order belief by asking: *Out of* 100 *possible counterparts, how many do you think will follow your messages (i.e., choose the number you mention in your message)?*

In the questionnaire we ask demographics (i.e., gender, age, education) and use Tanner et al. (2009)'s protected values survey similar as in Game A.

We have slightly adjusted the question supposed to elicit players' prosocial concerns:

Suppose the outcome from the roll of the 6-sided die is 5. To what extent do you judge the decision to send something other than Message 5: "The outcome from the roll of the 6-sided die is 5." as:

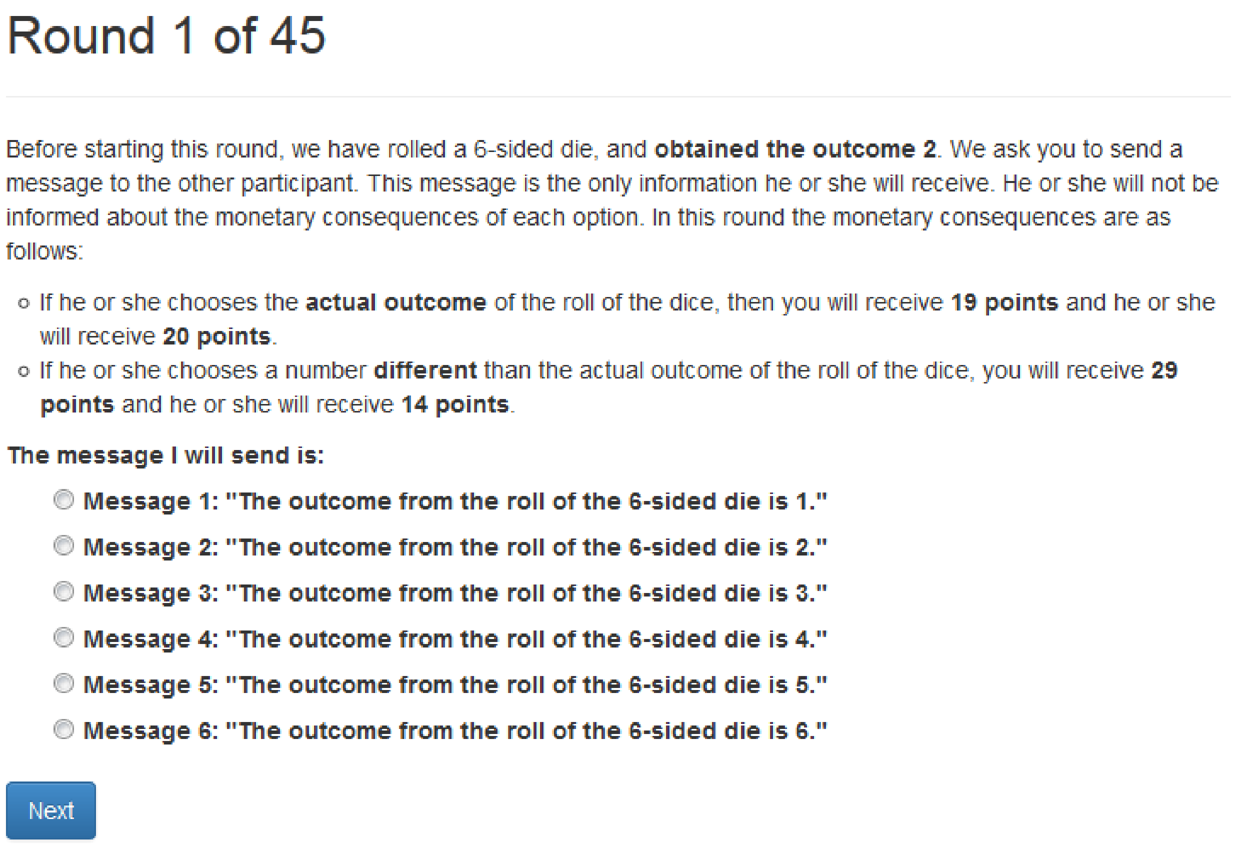


Figure 2.B.4: Decision situation for a Sender in Game B

- *1: Very dishonest ... 7: Very honest*
- *1: Very manipulative ... 7: Not at all manipulative*
- *1: Strongly associated with personal gains ... 7: Not at all associated with personal gains*
- *1: Strongly hurting the other participant ... 7: Not at all hurting the other participant*

Finally, the final screen (similar to Figure 2.B.2) of the experiment is shown before being returned to mTurk.[27]

[27] In fact, for a first pilot experiment, we have directly matched participants and shown them the variable payoffs they achieved. However, this caused, on the one hand, unsatisfactory waiting periods for some participants due to a delayed start of others, and, on the other hand, administrative effort for the experimenter due to players disconnecting in between.

2.B.2.2 Receiver

A participant who is assigned to be a Receiver in Game B is provided with the following instructions:

Instructions

This is a short experiment in decision making. Please read the instructions carefully. You may earn a considerable sum of money, depending on the decisions you make in the experiment.

You will be randomly matched with another (anonymous) participant in this experiment. Neither of you will know the identity of the other. The money you earn in this game will be paid to you in addition to a fixed amount of \$0.50 *for participation.*

We have rolled a 6-sided die, and told the outcome of it to the other participant, but we are not going to tell it to you.

After being informed of the roll of the die, the other participants sent a message to you. The message corresponds to a number from 1 to 6. There are six possible messages:

- *Message 1: "The outcome from the roll of the 6-sided die is 1."*
- *Message 2: "The outcome from the roll of the 6-sided die is 2."*
- *Message 3: "The outcome from the roll of the 6-sided die is 3."*
- *Message 4: "The outcome from the roll of the 6-sided die is 4."*
- *Message 5: "The outcome from the roll of the 6-sided die is 5."*
- *Message 6: "The outcome from the roll of the 6-sided die is 6."*

We will ask you to choose a number between 1 and 6. The message you received is the only information you will have regarding the roll of the die. Your choice of a number will determine the payments in the experiment according to two different options (Option X and Option Y), known only to the other participant.

If you will choose the same number as the number that came up in the roll of the die, both of you will be paid according to option A. If you will choose a number different than the actual number, you will both be paid according to option B.

For your convenience, these instructions will remain available to you on all subsequent screens of this study.

Right after the instructions, the Receiver's decision screen is shown asking for the Receiver's strategies conditional on all possible messages (see Figure 2.B.5):

Decisions

The other participant will send you a message. For every possible message we ask you to choose one of the six numbers. Your choice will determine the payments to both of you.

Decision 1: Suppose the other participant sent you: Message 1: "The outcome from the roll of the 6-sided die is 1." Given this message the number I choose is:

•1 • 2 • 3 • 4 • 5 • 6

Decision 2: Suppose the other participant sent you: Message 2: "The outcome from the roll of the 6-sided die is 2." Given this message the number I choose is:

•1 • 2 • 3 • 4 • 5 • 6

Decision 3: Suppose the other participant sent you: Message 3: "The outcome from the roll of the 6-sided die is 3." Given this message the number I choose is:

•1 • 2 • 3 • 4 • 5 • 6

Decision 4: Suppose the other participant sent you: Message 4: "The outcome from the roll of the 6-sided die is 4." Given this message the number I choose is:

•1 • 2 • 3 • 4 • 5 • 6

Decision 5: Suppose the other participant sent you: Message 5: "The outcome from the roll of the 6-sided die is 5." Given this message the number I choose is:

•1 • 2 • 3 • 4 • 5 • 6

Decision 6: Suppose the other participant sent you: Message 6: "The outcome from the roll of the 6-sided die is 6." Given this message the number I choose is:

•1 • 2 • 3 • 4 • 5 • 6

In a next step, we elicit the Receiver's first-order belief by asking: *Out of* 100 *possible counterparts, how many do you think will send you a truthful message (i.e., send the message with the number of the actual outcome from the roll of the 6-sided die)?*

Finally, a questionnaire and a final screen similar to the Sender's close the experiment.

Decisions

The other participant will send you a message. For every possible message we ask you to choose one of the six numbers. Your choice will determine the payments to both of you.

Decision 1: Suppose the other participant sent you: **Message 1: "The outcome from the roll of the 6-sided die is 1."**

Given this message the number I choose is:

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

Decision 2: Suppose the other participant sent you: **Message 2: "The outcome from the roll of the 6-sided die is 2."**

Given this message the number I choose is:

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

Decision 3: Suppose the other participant sent you: **Message 3: "The outcome from the roll of the 6-sided die is 3."**

Given this message the number I choose is:

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

Decision 4: Suppose the other participant sent you: **Message 4: "The outcome from the roll of the 6-sided die is 4."**

Given this message the number I choose is:

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

Decision 5: Suppose the other participant sent you: **Message 5: "The outcome from the roll of the 6-sided die is 5."**

Given this message the number I choose is:

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

Decision 6: Suppose the other participant sent you: **Message 6: "The outcome from the roll of the 6-sided die is 6."**

Given this message the number I choose is:

○ 1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6

Next

Figure 2.B.5: Decision situation for a Receiver in Game B

2.C Appendix C: Expected Utilities

In this section, we illustrate Senders' expected utility computation conditional on belonging to each of the five types.

2.C.1 Type EC

We know that EC Participants derive utility

$$U_i^{\mathrm{EC}}(a, \theta, m_i) = \Pi^S(\theta, a),$$

where $\Pi^S(\theta, a)$ indicates the Sender's monetary payoff in state θ when the Receiver plays action a.

As a consequence, the expected utility of lying equals

$$\mathbb{E}\left[U_i^{\mathrm{EC}}(a, \theta, \theta')\right] = \beta(\Pi + \Xi) + (1-\beta)(q\Pi + (1-q)(\Pi + \Xi)),$$

where $\Pi := \Pi^S(\theta, \theta)$ denotes the Sender's payoff if the Receiver matches the state ($a = \theta$).[28] Ξ indicates ECOST, β the Sender's first-order belief, and q the game-specific probability to achieve the low payoff if the Sender lies and the Receiver does not follow. In Game A, this probability equals $q = 1$, and in Game B $q = 1/5$. The expected utility of telling the truth is given by

$$\mathbb{E}\left[U_i^{\mathrm{EC}}(a, \theta, \theta)\right] = \beta\Pi + (1-\beta)(\Pi + \Xi).$$

In order to decide about lying or telling the truth, the Sender considers the difference

$$\mathbb{E}\left[U_i^{\mathrm{EC}}(a, \theta, \theta') - U_i^{\mathrm{EC}}(a, \theta, \theta)\right] = (\beta + \beta q - q)\,\Xi. \tag{2.C.1}$$

The greater Eq. (2.C.1), the higher is the likelihood that the Sender tells a lie conditional on being EC.

2.C.2 Type IA

We consider the following utility specification for inequity aversion

$$U_i^{\mathrm{IA}}(a, \theta, m_i) = \Pi^S(\theta, a) - \gamma_{IA} \max\left(\Pi^R(\theta, a) - \Pi^S(\theta, a), 0\right),$$

[28] Since incentives are misaligned, the Sender's payoff if $a = \theta$ always equals the lower of the two possible Sender payoffs.

where γ_{IA} denotes the Sender's distaste for negative inequity.

Therefore, we derive the Sender's expected utility of lying as

$$\mathbb{E}\left[U_i^{\mathrm{IA}}(a,\theta,\theta')\right] = \beta\left(\Pi + \Xi\right) + (1-\beta)\left(q(\Pi - \gamma_{IA}\delta) + (1-q)(\Pi+\Xi)\right),$$

where $\delta := \Pi^R(\theta,\theta) - \Pi^S(\theta,\theta)$ denotes negative inequity in case the Receiver matches the state. The expected utility of telling the truth equals

$$\mathbb{E}\left[U_i^{\mathrm{IA}}(a,\theta,\theta)\right] = \beta\left(\Pi - \gamma_{IA}\delta\right) + (1-\beta)\left(\Pi + \Xi\right),$$

implying a difference between lying and telling the truth of

$$\mathbb{E}\left[U_i^{\mathrm{IA}}(a,\theta,\theta') - U_i^{\mathrm{IA}}(a,\theta,\theta)\right] = \left(\beta + \beta q - q\right)\left(\Xi + \gamma_{IA}\delta\right).$$

2.C.3 Type RD

Reference-dependent preferences suggest that utility depends on personal expectations. We compute the utility specification as

$$U_i^{\mathrm{RD}}(a,\theta,m_i) = \Pi^S(\theta,a) + \gamma_{RD}\left(\Pi^S(\theta,a) - \mathbb{E}\left[\Pi^S(\theta,a)\right]\right),$$

where γ_{RD} reflects a coefficient which measures the importance of the reference point.

We derive an expected utility of lying

$$\begin{aligned}\mathbb{E}\left[U_i^{\mathrm{RD}}(a,\theta,\theta')\right] = \beta\Big[&(1+\gamma_{RD})\left(\Pi+\Xi\right) - \gamma_{RD}\mathbb{E}\left[\Pi^S(\theta,a)\right]\Big]\\ &+ (1-\beta)\Big[q\left((1+\gamma_{RD})\Pi - \gamma_{RD}\mathbb{E}\left[\Pi^S(\theta,a)\right]\right)\\ &+ (1-q)(1+\gamma_{RD})\left(\Pi+\Xi\right) - \gamma_{RD}\mathbb{E}\left[\Pi^S(\theta,a)\right]\Big],\end{aligned}$$

and an expected utility of telling the truth

$$\begin{aligned}\mathbb{E}\left[U_i^{\mathrm{RD}}(a,\theta,\theta)\right] = \beta\Big[&(1+\gamma_{RD})\Pi - \gamma_{RD}\mathbb{E}\left[\Pi^S(\theta,a)\right]\Big]\\ &+ (1-\beta)\Big[(1+\gamma_{RD})\left(\Pi+\Xi\right) - \gamma_{RD}\mathbb{E}\left[\Pi^S(\theta,a)\right]\Big].\end{aligned}$$

The resulting difference equals

$$\mathbb{E}\left[U_i^{\mathrm{RD}}(a,\theta,\theta') - U_i^{\mathrm{RD}}(a,\theta,\theta)\right] = \left(\beta+\beta q - q\right)\left(1+\gamma_{RD}\right)\Xi.$$

2.C.4 Type LC

Lying costs refer to the idea that participants incur a private cost when telling a lie. We assume a utility representation

$$U_i^{\mathrm{LC}}(a,\theta,m_i) = \Pi^S(\theta,a) - \left(\alpha + \beta\Xi + \rho\Xi^2\right)\mathbb{1}_{m_i\neq\theta},$$

where α, β, and ρ are the coefficients for the constant, the linear and the convex term of ECOST, respectively. Lying costs occur if the Sender lies, i.e., if the message deviates from the state.

We derive the expected utility of lying to be given by

$$\mathbb{E}\left[U_i^{\mathrm{LC}}(a,\theta,\theta')\right] = \beta\left(\Pi + \Xi - \Psi\right) + (1-\beta)\left(q(\Pi - \Psi) + (1-q)\left(\Pi + \Xi - \Psi\right)\right),$$

where $\Psi := (a + b\Xi + c\Xi^2)$ denotes the Sender's cost of lying. The expected utility of telling the truth equals

$$\mathbb{E}\left[U_i^{\mathrm{LC}}(a,\theta,\theta)\right] = \beta\left(\Pi + \Xi\right) + (1-\beta)\left(\Pi + \Xi\right),$$

respectively, implying the marginal utility increase of lying to be

$$\mathbb{E}\left[U_i^{\mathrm{LC}}(a,\theta,\theta') - U_i^{\mathrm{LC}}(a,\theta,\theta)\right] = (\beta + \beta q - q)\,\Xi - \Psi.$$

2.C.5 Type SN

The idea of reputation for honesty is that participants dislike to act in the same way as dishonest people. We apply the utility characterization

$$U_i^{\mathrm{SN}}(a,\theta,m) = \Pi^S(\theta,a) - \gamma_{SN}\sum_{j\neq i}\frac{\mathbb{1}_{m_j=\theta}}{N}\mathbb{1}_{m_i\neq\theta},\ j\in\{1,\dots,N\},$$

where γ_{SN} is the individual coefficient that loads on the average proportion of truthful reports in a given game. m denotes the entire set of messages.

Having this utility representation, the expected utility of lying equals

$$\mathbb{E}\left[U_i^{\mathrm{SN}}(a,\theta,m)\right] = \beta\left(\Pi + \Xi - \gamma_{SN}\overline{m}\right) + (1-\beta)\left(q\left(\Pi - \gamma_{SN}\overline{m}\right) + (1-q)\left(\Pi + \Xi - \gamma_{SN}\overline{m}\right)\right),$$

where $\overline{m} := \sum_{j\neq i}\frac{\mathbb{1}_{m_j=\theta}}{N}$ denotes the average number of honest messages. The corresponding expected utility of telling the truth is given by

$$\mathbb{E}\left[U_i^{\mathrm{SN}}(a,\theta,m)\right] = \beta\Pi + (1-\beta)\left(\Pi + \Xi\right),$$

and the resulting difference between lying and telling the truth is computed as

$$\mathbb{E}\left[U_i^{\mathrm{SN}}(a,\theta,m) - U_i^{\mathrm{SN}}(a,\theta,m)\right] = (\beta + \beta q - q)\,\Xi - \gamma_{SN}\overline{m}.$$

2.D Appendix D: Additional Tables

Table 2.D.1: List of all worlds

	Number of types	Available types
1	one type	EC
2		IA
3		RD
4		LC
5		SN
6	two types	EC & IA
7		EC & RD
8		EC & LC
9		EC & SN
10		IA & RD
11		IA & LC
12		IA & SN
13		RD & LC
14		RD & SN
15		LC & SN
16	three types	EC, IA & RD
17		EC, IA & LC
18		EC, IA & SN
19		EC, RD & LC
20		EC, RD & SN
21		EC, LC & SN
22		IA, RD & LC
23		IA, RD & SN
24		IA, LC & SN
25		RD, LC & SN
26	four types	EC, IA, RD & LC
27		EC, IA, RD & SN
28		IA, RD, LC & SN
29		EC, RD, LC & SN
30		EC, IA, LC & SN
31	five types	EC, IA, RD, LC & SN

This table presents a list of all 31 worlds. EC reflects the standard economic model, IA inequity aversion, RD reference-dependent preferences, LC lying costs, and SN reputation or social norms.

Table 2.D.2: In- and out-of-sample accuracy - Game A

			Decisions used to estimate and classify																					
			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)
1-type	EC	in	0.521	0.522	0.525	0.524	0.524	0.525	0.524	0.524	0.524	0.525	0.525	0.525	0.525	0.525	0.525	0.525	0.525	0.525	0.525	0.526	0.526	0.525
		out	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526
	IA	in	0.938	0.929	0.910	0.892	0.882	0.873	0.866	0.861	0.858	0.855	0.853	0.851	0.849	0.848	0.847	0.846	0.845	0.844	0.843	0.843	0.842	0.841
		out	0.722	0.774	0.790	0.797	0.802	0.805	0.807	0.809	0.811	0.812	0.813	0.814	0.815	0.816	0.816	0.816	0.817	0.817	0.818	0.818	0.818	0.819
	RD	in	0.897	0.866	0.858	0.854	0.849	0.844	0.840	0.837	0.836	0.834	0.833	0.832	0.831	0.830	0.830	0.829	0.828	0.828	0.828	0.827	0.827	0.826
		out	0.699	0.768	0.791	0.799	0.803	0.806	0.808	0.809	0.810	0.811	0.811	0.812	0.813	0.813	0.813	0.814	0.814	0.814	0.814	0.814	0.815	0.815
	LC	in	1.000	0.997	0.991	0.976	0.961	0.949	0.939	0.930	0.923	0.918	0.913	0.910	0.906	0.904	0.902	0.899	0.898	0.896	0.895	0.894	0.893	0.892
		out	0.772	0.794	0.805	0.812	0.817	0.821	0.824	0.828	0.831	0.834	0.836	0.839	0.841	0.843	0.845	0.847	0.848	0.850	0.852	0.853	0.854	0.855
	SN	in	1.000	0.971	0.957	0.939	0.930	0.922	0.917	0.912	0.909	0.906	0.904	0.902	0.900	0.899	0.898	0.896	0.896	0.895	0.894	0.893	0.893	0.892
		out	0.771	0.804	0.827	0.837	0.844	0.849	0.854	0.857	0.859	0.861	0.863	0.864	0.866	0.867	0.867	0.868	0.869	0.870	0.871	0.871	0.872	0.872
2-types	EC, IA	in	0.614	0.633	0.895	0.883	0.877	0.869	0.865	0.860	0.857	0.855	0.853	0.851	0.849	0.848	0.847	0.846	0.845	0.844	0.843	0.843	0.842	0.841
		out	0.496	0.525	0.793	0.800	0.805	0.807	0.810	0.812	0.813	0.814	0.815	0.815	0.816	0.817	0.817	0.817	0.817	0.818	0.818	0.819	0.819	0.819
	EC, RD	in	0.576	0.547	0.791	0.820	0.825	0.832	0.830	0.831	0.831	0.833	0.833	0.832	0.831	0.830	0.830	0.829	0.829	0.828	0.828	0.828	0.827	0.827
		out	0.507	0.519	0.772	0.801	0.805	0.812	0.813	0.814	0.814	0.816	0.816	0.816	0.816	0.816	0.816	0.816	0.817	0.817	0.817	0.817	0.815	0.815
	EC, LC	in	1.000	0.997	0.989	0.962	0.935	0.925	0.919	0.914	0.909	0.905	0.902	0.902	0.895	0.893	0.891	0.886	0.887	0.883	0.883	0.881	0.884	0.886
		out	0.772	0.794	0.807	0.824	0.833	0.837	0.839	0.840	0.843	0.845	0.846	0.847	0.848	0.850	0.851	0.850	0.853	0.851	0.853	0.852	0.855	0.858
	EC, SN	in	1.000	0.971	0.957	0.939	0.930	0.922	0.917	0.912	0.909	0.906	0.904	0.902	0.900	0.899	0.898	0.896	0.896	0.895	0.894	0.893	0.893	0.892
		out	0.771	0.804	0.827	0.837	0.844	0.849	0.854	0.857	0.859	0.861	0.863	0.864	0.866	0.867	0.867	0.868	0.869	0.870	0.871	0.871	0.872	0.872
	IA, RD	in	0.939	0.924	0.910	0.891	0.881	0.872	0.866	0.861	0.858	0.856	0.854	0.852	0.850	0.848	0.847	0.846	0.845	0.845	0.844	0.843	0.843	0.842
		out	0.720	0.772	0.790	0.797	0.803	0.805	0.809	0.811	0.812	0.814	0.815	0.816	0.817	0.818	0.818	0.819	0.820	0.820	0.820	0.821	0.821	0.821
	IA, LC	in	1.000	0.997	0.982	0.952	0.935	0.931	0.922	0.915	0.911	0.908	0.902	0.899	0.896	0.894	0.892	0.888	0.887	0.884	0.883	0.882	0.882	0.883
		out	0.771	0.794	0.802	0.819	0.822	0.829	0.834	0.837	0.840	0.842	0.843	0.845	0.847	0.849	0.850	0.849	0.850	0.849	0.850	0.851	0.852	0.855
	IA, SN	in	1.000	0.971	0.957	0.940	0.929	0.920	0.914	0.910	0.905	0.903	0.901	0.898	0.896	0.895	0.894	0.893	0.892	0.892	0.891	0.890	0.890	0.890
		out	0.771	0.804	0.827	0.837	0.845	0.850	0.855	0.858	0.861	0.863	0.865	0.866	0.867	0.868	0.869	0.870	0.870	0.871	0.872	0.872	0.873	0.873
	RD, LC	in	1.000	0.997	0.989	0.971	0.952	0.934	0.925	0.915	0.910	0.906	0.902	0.902	0.895	0.893	0.892	0.886	0.887	0.883	0.883	0.881	0.884	0.886
		out	0.772	0.794	0.807	0.817	0.823	0.829	0.834	0.837	0.840	0.842	0.844	0.846	0.847	0.849	0.850	0.849	0.852	0.850	0.852	0.852	0.855	0.857
	RD, SN	in	1.000	0.971	0.957	0.939	0.930	0.922	0.917	0.912	0.909	0.906	0.904	0.902	0.900	0.899	0.898	0.896	0.896	0.895	0.894	0.893	0.893	0.892
		out	0.771	0.804	0.827	0.837	0.844	0.849	0.854	0.857	0.860	0.861	0.863	0.865	0.866	0.867	0.868	0.869	0.869	0.870	0.871	0.872	0.872	0.872
	LC, SN	in	1.000	0.997	0.980	0.962	0.946	0.938	0.930	0.925	0.920	0.917	0.913	0.911	0.907	0.906	0.904	0.902	0.901	0.900	0.899	0.899	0.899	0.898
		out	0.766	0.793	0.818	0.832	0.839	0.841	0.847	0.850	0.852	0.853	0.855	0.857	0.860	0.860	0.863	0.864	0.865	0.866	0.867	0.867	0.868	0.869
3-types	EC, IA, RD	in	0.625	0.629	0.895	0.883	0.878	0.869	0.865	0.861	0.858	0.855	0.853	0.852	0.850	0.848	0.848	0.846	0.845	0.845	0.844	0.843	0.843	0.842
		out	0.496	0.523	0.793	0.800	0.805	0.808	0.811	0.813	0.814	0.816	0.816	0.817	0.818	0.819	0.819	0.819	0.820	0.820	0.820	0.821	0.821	0.821
	EC, IA, LC	in	1.000	0.997	0.981	0.950	0.931	0.928	0.921	0.914	0.910	0.907	0.902	0.899	0.896	0.894	0.892	0.888	0.887	0.884	0.883	0.882	0.882	0.883
		out	0.771	0.794	0.802	0.821	0.824	0.832	0.835	0.838	0.841	0.843	0.844	0.846	0.848	0.849	0.850	0.849	0.851	0.850	0.851	0.852	0.852	0.855
	EC, IA, SN	in	1.000	0.971	0.957	0.940	0.929	0.920	0.914	0.910	0.905	0.903	0.901	0.898	0.896	0.895	0.894	0.893	0.892	0.892	0.891	0.890	0.890	0.890
		out	0.771	0.804	0.827	0.837	0.845	0.850	0.855	0.858	0.861	0.863	0.865	0.866	0.867	0.868	0.869	0.870	0.870	0.871	0.872	0.872	0.873	0.873
	EC, RD, LC	in	1.000	0.997	0.989	0.962	0.935	0.925	0.919	0.914	0.909	0.905	0.902	0.902	0.895	0.894	0.891	0.887	0.888	0.884	0.886	0.883	0.885	0.888
		out	0.772	0.794	0.808	0.823	0.833	0.836	0.839	0.840	0.843	0.845	0.846	0.847	0.848	0.850	0.851	0.851	0.853	0.851	0.853	0.852	0.855	0.857
	EC, RD, SN	in	1.000	0.971	0.957	0.939	0.930	0.922	0.917	0.912	0.909	0.906	0.904	0.902	0.900	0.899	0.898	0.896	0.896	0.895	0.894	0.893	0.893	0.892
		out	0.771	0.804	0.827	0.837	0.844	0.849	0.854	0.857	0.860	0.861	0.863	0.865	0.866	0.867	0.868	0.869	0.869	0.870	0.871	0.872	0.872	0.872
	EC, LC, SN	in	1.000	0.997	0.980	0.962	0.946	0.938	0.930	0.925	0.920	0.917	0.913	0.911	0.907	0.906	0.904	0.902	0.901	0.900	0.899	0.899	0.899	0.898
		out	0.766	0.793	0.818	0.832	0.839	0.841	0.847	0.850	0.852	0.853	0.855	0.857	0.860	0.860	0.863	0.864	0.865	0.866	0.867	0.867	0.868	0.869
	IA, RD, LC	in	1.000	0.997	0.981	0.952	0.935	0.930	0.922	0.915	0.911	0.907	0.902	0.899	0.896	0.894	0.892	0.888	0.887	0.884	0.883	0.882	0.882	0.883
		out	0.771	0.794	0.803	0.819	0.822	0.830	0.834	0.837	0.841	0.843	0.844	0.846	0.848	0.849	0.850	0.849	0.851	0.850	0.851	0.852	0.852	0.855
	IA, RD, SN	in	1.000	0.971	0.957	0.940	0.929	0.920	0.914	0.910	0.905	0.903	0.901	0.898	0.896	0.895	0.894	0.893	0.892	0.891	0.891	0.890	0.890	0.889
		out	0.771	0.804	0.827	0.837	0.845	0.850	0.855	0.858	0.861	0.863	0.865	0.866	0.867	0.868	0.869	0.870	0.871	0.872	0.872	0.873	0.873	0.874
	IA, LC, SN	in	1.000	0.997	0.980	0.962	0.944	0.936	0.927	0.923	0.917	0.914	0.911	0.907	0.904	0.902	0.900	0.899	0.898	0.897	0.896	0.896	0.896	0.895
		out	0.766	0.793	0.817	0.832	0.840	0.843	0.849	0.852	0.854	0.855	0.857	0.858	0.861	0.862	0.864	0.865	0.866	0.867	0.868	0.868	0.869	0.870
	RD, LC, SN	in	1.000	0.997	0.980	0.962	0.946	0.938	0.930	0.925	0.920	0.917	0.913	0.911	0.907	0.906	0.904	0.902	0.901	0.900	0.899	0.899	0.899	0.898
		out	0.766	0.793	0.819	0.832	0.839	0.841	0.847	0.850	0.852	0.853	0.855	0.857	0.860	0.860	0.863	0.864	0.865	0.866	0.867	0.867	0.868	0.869
4-types	EC, IA, RD, LC	in	1.000	0.997	0.981	0.950	0.931	0.928	0.921	0.914	0.910	0.907	0.902	0.899	0.896	0.894	0.892	0.888	0.887	0.884	0.883	0.882	0.882	0.883
		out	0.771	0.794	0.803	0.821	0.824	0.832	0.835	0.838	0.841	0.843	0.844	0.846	0.848	0.849	0.850	0.849	0.851	0.850	0.851	0.852	0.852	0.855
	EC, IA, RD, SN	in	1.000	0.971	0.957	0.940	0.929	0.920	0.914	0.910	0.905	0.903	0.901	0.898	0.896	0.895	0.894	0.893	0.892	0.891	0.891	0.890	0.890	0.889
		out	0.771	0.804	0.827	0.837	0.845	0.850	0.855	0.858	0.861	0.863	0.865	0.866	0.867	0.868	0.869	0.870	0.871	0.872	0.872	0.873	0.873	0.874
	IA, RD, LC, SN	in	1.000	0.997	0.980	0.962	0.944	0.936	0.927	0.923	0.917	0.914	0.911	0.907	0.904	0.902	0.900	0.899	0.898	0.897	0.896	0.896	0.896	0.895
		out	0.766	0.793	0.818	0.832	0.840	0.843	0.849	0.852	0.854	0.855	0.857	0.858	0.861	0.862	0.864	0.865	0.866	0.867	0.868	0.868	0.869	0.870
	EC, RD, LC, SN	in	1.000	0.997	0.980	0.962	0.946	0.938	0.930	0.925	0.920	0.917	0.913	0.911	0.907	0.906	0.904	0.902	0.901	0.900	0.899	0.899	0.899	0.898
		out	0.766	0.793	0.819	0.832	0.839	0.841	0.847	0.850	0.852	0.853	0.855	0.857	0.860	0.860	0.863	0.864	0.865	0.866	0.867	0.867	0.868	0.869
	EC, IA, LC, SN	in	1.000	0.997	0.980	0.962	0.944	0.936	0.927	0.923	0.917	0.914	0.911	0.907	0.904	0.902	0.900	0.899	0.898	0.897	0.896	0.896	0.896	0.895
		out	0.766	0.793	0.817	0.832	0.840	0.843	0.849	0.852	0.854	0.855	0.857	0.858	0.861	0.862	0.864	0.865	0.866	0.867	0.868	0.868	0.869	0.870
5-types	EC, IA, RD, LC, SN	in	1.000	0.997	0.980	0.962	0.944	0.936	0.927	0.923	0.917	0.914	0.911	0.907	0.904	0.902	0.900	0.899	0.898	0.897	0.896	0.896	0.896	0.895
		out	0.766	0.793	0.818	0.832	0.840	0.843	0.849	0.852	0.854	0.855	0.857	0.858	0.861	0.862	0.864	0.865	0.866	0.867	0.868	0.868	0.869	0.870

Table 2.D.2: In- and out-of-sample accuracy - Game A (continued)

			Decisions used to estimate and classify																					
			(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)	(31)	(32)	(33)	(34)	(35)	(36)	(37)	(38)	(39)	(40)	(41)	(42)	(43)	(44)
1-type	EC	in	0.525	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.526	0.525	0.525	0.526	0.526	0.526	0.526	0.526	0.526
		out	0.526	0.526	0.526	0.526	0.525	0.525	0.525	0.526	0.526	0.526	0.526	0.525	0.525	0.526	0.526	0.526	0.526	0.525	0.524	0.524	0.522	0.521
	IA	in	0.840	0.840	0.840	0.839	0.839	0.839	0.838	0.838	0.837	0.837	0.837	0.837	0.836	0.836	0.836	0.835	0.835	0.835	0.835	0.835	0.835	0.834
		out	0.819	0.819	0.819	0.819	0.819	0.820	0.820	0.820	0.820	0.821	0.821	0.821	0.821	0.822	0.823	0.823	0.823	0.823	0.823	0.824	0.824	0.824
	RD	in	0.826	0.826	0.826	0.826	0.826	0.825	0.825	0.825	0.825	0.825	0.825	0.825	0.824	0.824	0.824	0.824	0.824	0.824	0.824	0.824	0.824	0.824
		out	0.815	0.815	0.815	0.815	0.815	0.815	0.815	0.816	0.815	0.816	0.816	0.815	0.816	0.816	0.817	0.818	0.818	0.817	0.817	0.817	0.816	0.812
	LC	in	0.891	0.891	0.890	0.890	0.889	0.889	0.888	0.888	0.887	0.887	0.886	0.886	0.886	0.886	0.885	0.885	0.885	0.885	0.884	0.884	0.884	0.884
		out	0.857	0.858	0.859	0.859	0.860	0.860	0.861	0.862	0.862	0.863	0.863	0.864	0.865	0.865	0.866	0.866	0.867	0.866	0.867	0.867	0.867	0.869
	SN	in	0.892	0.891	0.891	0.891	0.890	0.890	0.890	0.889	0.889	0.889	0.888	0.888	0.888	0.888	0.887	0.887	0.887	0.887	0.887	0.886	0.886	0.886
		out	0.872	0.873	0.873	0.873	0.873	0.873	0.874	0.874	0.874	0.874	0.875	0.875	0.875	0.875	0.875	0.875	0.876	0.876	0.876	0.876	0.876	0.877
2-types	EC, IA	in	0.840	0.840	0.840	0.839	0.839	0.839	0.838	0.838	0.837	0.837	0.837	0.837	0.836	0.836	0.836	0.835	0.835	0.835	0.835	0.835	0.835	0.834
		out	0.819	0.819	0.819	0.819	0.819	0.820	0.820	0.820	0.820	0.821	0.821	0.821	0.821	0.822	0.823	0.823	0.823	0.823	0.823	0.824	0.824	0.824
	EC, RD	in	0.826	0.826	0.826	0.826	0.826	0.825	0.825	0.825	0.825	0.825	0.825	0.825	0.824	0.824	0.824	0.824	0.824	0.824	0.824	0.824	0.824	0.824
		out	0.815	0.815	0.815	0.815	0.815	0.815	0.815	0.816	0.815	0.816	0.816	0.815	0.816	0.816	0.817	0.818	0.818	0.817	0.817	0.817	0.816	0.812
	EC, LC	in	0.883	0.883	0.885	0.885	0.884	0.884	0.883	0.883	0.883	0.885	0.885	0.885	0.885	0.885	0.885	0.885	0.885	0.884	0.883	0.883	0.883	0.883
		out	0.857	0.857	0.860	0.860	0.860	0.861	0.861	0.861	0.862	0.865	0.865	0.865	0.867	0.867	0.868	0.868	0.869	0.867	0.869	0.868	0.868	0.868
	EC, SN	in	0.892	0.891	0.891	0.891	0.890	0.890	0.890	0.889	0.889	0.889	0.888	0.888	0.888	0.888	0.887	0.887	0.887	0.887	0.887	0.886	0.886	0.886
		out	0.872	0.873	0.873	0.873	0.873	0.873	0.874	0.874	0.874	0.874	0.875	0.875	0.875	0.875	0.875	0.875	0.876	0.876	0.876	0.876	0.876	0.877
	IA, RD	in	0.841	0.841	0.840	0.840	0.839	0.839	0.839	0.838	0.838	0.837	0.837	0.837	0.836	0.836	0.836	0.835	0.835	0.835	0.835	0.835	0.835	0.834
		out	0.821	0.821	0.821	0.821	0.821	0.822	0.822	0.822	0.822	0.822	0.823	0.822	0.823	0.823	0.824	0.824	0.825	0.824	0.824	0.825	0.824	0.824
	IA, LC	in	0.881	0.881	0.884	0.884	0.881	0.883	0.882	0.882	0.882	0.884	0.884	0.883	0.884	0.884	0.884	0.883	0.883	0.882	0.881	0.881	0.881	0.881
		out	0.854	0.855	0.859	0.859	0.858	0.860	0.861	0.861	0.861	0.865	0.865	0.865	0.867	0.867	0.868	0.868	0.868	0.867	0.868	0.868	0.867	0.868
	IA, SN	in	0.889	0.889	0.888	0.888	0.888	0.888	0.887	0.887	0.887	0.887	0.886	0.886	0.886	0.886	0.885	0.885	0.885	0.885	0.884	0.884	0.884	0.884
		out	0.873	0.874	0.874	0.874	0.874	0.874	0.875	0.875	0.875	0.875	0.875	0.875	0.876	0.876	0.876	0.876	0.876	0.876	0.876	0.877	0.876	0.877
	RD, LC	in	0.883	0.883	0.885	0.885	0.884	0.884	0.883	0.883	0.883	0.885	0.885	0.885	0.885	0.885	0.885	0.885	0.885	0.883	0.883	0.883	0.883	0.883
		out	0.856	0.856	0.859	0.859	0.860	0.860	0.861	0.861	0.862	0.864	0.865	0.865	0.866	0.867	0.867	0.868	0.868	0.867	0.868	0.868	0.867	0.868
	RD, SN	in	0.892	0.891	0.891	0.891	0.890	0.890	0.890	0.889	0.889	0.889	0.888	0.888	0.888	0.888	0.887	0.887	0.887	0.887	0.887	0.886	0.886	0.886
		out	0.873	0.873	0.874	0.874	0.874	0.874	0.875	0.875	0.875	0.875	0.876	0.875	0.876	0.876	0.876	0.877	0.877	0.877	0.877	0.877	0.877	0.877
	LC, SN	in	0.897	0.897	0.896	0.896	0.896	0.895	0.895	0.894	0.894	0.894	0.893	0.893	0.893	0.892	0.892	0.892	0.892	0.892	0.891	0.891	0.890	0.891
		out	0.870	0.870	0.871	0.871	0.871	0.872	0.872	0.873	0.873	0.873	0.874	0.874	0.874	0.875	0.875	0.875	0.876	0.876	0.876	0.876	0.876	0.877
3-types	EC, IA, RD	in	0.841	0.841	0.840	0.840	0.839	0.839	0.839	0.838	0.838	0.837	0.837	0.837	0.836	0.836	0.836	0.835	0.835	0.835	0.835	0.835	0.835	0.834
		out	0.821	0.821	0.821	0.821	0.821	0.822	0.822	0.822	0.822	0.822	0.823	0.822	0.823	0.823	0.824	0.824	0.825	0.824	0.824	0.825	0.824	0.824
	EC, IA, LC	in	0.881	0.881	0.884	0.884	0.881	0.883	0.882	0.882	0.882	0.884	0.884	0.883	0.884	0.884	0.884	0.883	0.883	0.882	0.881	0.881	0.881	0.881
		out	0.854	0.855	0.859	0.859	0.858	0.860	0.861	0.861	0.861	0.865	0.865	0.865	0.867	0.867	0.868	0.868	0.868	0.867	0.868	0.868	0.867	0.868
	EC, IA, SN	in	0.889	0.889	0.888	0.888	0.888	0.888	0.887	0.887	0.887	0.887	0.886	0.886	0.886	0.886	0.885	0.885	0.885	0.885	0.884	0.884	0.884	0.884
		out	0.873	0.874	0.874	0.874	0.874	0.874	0.875	0.875	0.875	0.875	0.875	0.875	0.876	0.876	0.876	0.876	0.876	0.876	0.876	0.877	0.876	0.877
	EC, RD, LC	in	0.884	0.885	0.887	0.886	0.886	0.885	0.887	0.887	0.884	0.885	0.887	0.885	0.885	0.885	0.886	0.885	0.885	0.883	0.883	0.883	0.884	0.883
		out	0.857	0.856	0.859	0.859	0.860	0.860	0.862	0.863	0.862	0.864	0.865	0.865	0.867	0.867	0.868	0.868	0.868	0.866	0.868	0.868	0.868	0.869
	EC, RD, SN	in	0.892	0.891	0.891	0.891	0.890	0.890	0.890	0.889	0.889	0.889	0.888	0.888	0.888	0.888	0.887	0.887	0.887	0.887	0.887	0.886	0.886	0.886
		out	0.873	0.873	0.874	0.874	0.874	0.874	0.875	0.875	0.875	0.875	0.876	0.875	0.876	0.876	0.876	0.877	0.877	0.877	0.877	0.877	0.877	0.877
	EC, LC, SN	in	0.897	0.897	0.896	0.896	0.896	0.895	0.895	0.894	0.894	0.894	0.893	0.893	0.893	0.892	0.892	0.892	0.892	0.892	0.891	0.891	0.890	0.891
		out	0.870	0.870	0.871	0.871	0.871	0.872	0.872	0.873	0.873	0.873	0.874	0.874	0.874	0.875	0.875	0.875	0.876	0.876	0.876	0.876	0.876	0.877
	IA, RD, LC	in	0.881	0.881	0.884	0.884	0.881	0.883	0.882	0.882	0.882	0.884	0.884	0.883	0.884	0.884	0.884	0.883	0.883	0.882	0.881	0.881	0.881	0.881
		out	0.854	0.855	0.859	0.859	0.858	0.860	0.861	0.861	0.861	0.865	0.865	0.865	0.867	0.867	0.868	0.868	0.868	0.867	0.868	0.868	0.867	0.868
	IA, RD, SN	in	0.889	0.889	0.888	0.888	0.888	0.888	0.887	0.887	0.887	0.887	0.886	0.886	0.886	0.886	0.885	0.885	0.885	0.885	0.885	0.884	0.884	0.884
		out	0.874	0.874	0.875	0.875	0.875	0.875	0.875	0.875	0.875	0.876	0.876	0.876	0.877	0.876	0.877	0.877	0.877	0.877	0.877	0.877	0.877	0.877
	IA, LC, SN	in	0.894	0.894	0.893	0.893	0.893	0.893	0.892	0.892	0.892	0.891	0.891	0.891	0.890	0.890	0.890	0.890	0.889	0.889	0.888	0.889	0.888	0.888
		out	0.871	0.871	0.872	0.872	0.872	0.873	0.873	0.873	0.874	0.874	0.874	0.874	0.875	0.875	0.876	0.876	0.876	0.875	0.876	0.876	0.875	0.877
	RD, LC, SN	in	0.897	0.897	0.896	0.896	0.896	0.895	0.895	0.894	0.894	0.894	0.893	0.893	0.893	0.892	0.892	0.892	0.892	0.892	0.891	0.891	0.890	0.891
		out	0.870	0.870	0.871	0.871	0.871	0.872	0.872	0.873	0.873	0.873	0.874	0.874	0.874	0.875	0.875	0.875	0.876	0.876	0.876	0.876	0.876	0.877
4-types	EC, IA, RD, LC	in	0.881	0.881	0.884	0.884	0.881	0.883	0.882	0.882	0.882	0.884	0.884	0.883	0.884	0.884	0.884	0.883	0.883	0.882	0.881	0.881	0.881	0.881
		out	0.854	0.855	0.859	0.859	0.858	0.860	0.861	0.861	0.861	0.865	0.865	0.865	0.867	0.867	0.868	0.868	0.868	0.867	0.868	0.868	0.867	0.868
	EC, IA, RD, SN	in	0.889	0.889	0.888	0.888	0.888	0.888	0.887	0.887	0.887	0.887	0.886	0.886	0.886	0.886	0.885	0.885	0.885	0.885	0.885	0.884	0.884	0.884
		out	0.874	0.874	0.875	0.875	0.875	0.875	0.875	0.875	0.875	0.876	0.876	0.876	0.877	0.876	0.877	0.877	0.877	0.877	0.877	0.877	0.877	0.877
	IA, RD, LC, SN	in	0.894	0.894	0.893	0.893	0.893	0.893	0.892	0.892	0.892	0.891	0.891	0.891	0.890	0.890	0.890	0.890	0.889	0.889	0.888	0.889	0.888	0.888
		out	0.871	0.871	0.872	0.872	0.872	0.873	0.873	0.873	0.874	0.874	0.874	0.874	0.875	0.875	0.876	0.876	0.876	0.875	0.876	0.876	0.875	0.877
	EC, RD, LC, SN	in	0.897	0.897	0.896	0.896	0.896	0.895	0.895	0.894	0.894	0.894	0.893	0.893	0.893	0.892	0.892	0.892	0.892	0.892	0.891	0.891	0.890	0.891
		out	0.870	0.870	0.871	0.871	0.871	0.872	0.872	0.873	0.873	0.873	0.874	0.874	0.874	0.875	0.875	0.875	0.876	0.876	0.876	0.876	0.876	0.877
	EC, IA, LC, SN	in	0.894	0.894	0.893	0.893	0.893	0.893	0.892	0.892	0.892	0.891	0.891	0.891	0.890	0.890	0.890	0.890	0.889	0.889	0.888	0.889	0.888	0.888
		out	0.871	0.871	0.872	0.872	0.872	0.873	0.873	0.873	0.874	0.874	0.874	0.874	0.875	0.875	0.876	0.876	0.876	0.875	0.876	0.876	0.875	0.877
5-types	EC, IA, RD, LC, SN	in	0.894	0.894	0.893	0.893	0.893	0.893	0.892	0.892	0.892	0.891	0.891	0.891	0.890	0.890	0.890	0.890	0.889	0.889	0.888	0.889	0.888	0.888
		out	0.871	0.871	0.872	0.872	0.872	0.873	0.873	0.873	0.874	0.874	0.874	0.874	0.875	0.875	0.876	0.876	0.876	0.875	0.876	0.876	0.875	0.877

This table presents the accuracies of the in-sample and out-of-sample fits of all 31 worlds in Game A. The accuracy is derived as the ratio of correctly fitted decisions over the total number of decisions, whereas $M = 1000$ random selections were used. We assume Senders' first-order beliefs to equal one. EC reflects the standard economic model, IA inequity aversion, RD reference-dependent preferences, LC lying costs, and SN reputation or social norms. Each column corresponds to a specific number of decisions used to estimate individual parameters, classify participants into types and to compute the in-sample performance. The remaining decisions are used to derive the out-of-sample fit.

Table 2.D.3: In- and out-of-sample accuracy standard errors - Game A

			Decisions used to estimate and classify																					
			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)
1-type	EC	in	1.722	0.258	0.199	0.169	0.153	0.137	0.126	0.115	0.106	0.100	0.093	0.087	0.083	0.078	0.075	0.073	0.069	0.067	0.063	0.060	0.058	0.055
		out	0.039	0.012	0.014	0.016	0.019	0.021	0.023	0.025	0.027	0.029	0.030	0.032	0.034	0.035	0.037	0.040	0.042	0.044	0.046	0.048	0.050	0.053
	IA	in	2.029	0.210	0.168	0.138	0.121	0.110	0.100	0.091	0.085	0.080	0.075	0.071	0.067	0.062	0.060	0.058	0.055	0.052	0.051	0.049	0.046	0.044
		out	1.475	0.161	0.090	0.048	0.043	0.035	0.033	0.031	0.029	0.029	0.030	0.031	0.032	0.033	0.034	0.036	0.037	0.039	0.040	0.041	0.043	0.045
	RD	in	2.527	0.235	0.175	0.124	0.103	0.095	0.089	0.080	0.074	0.070	0.066	0.062	0.059	0.056	0.053	0.052	0.050	0.048	0.045	0.043	0.041	0.040
		out	1.672	0.218	0.126	0.070	0.035	0.032	0.030	0.030	0.031	0.032	0.034	0.035	0.035	0.037	0.039	0.041	0.042	0.044	0.045	0.047	0.049	0.050
	LC	in	0.000	0.050	0.069	0.068	0.060	0.056	0.055	0.052	0.050	0.047	0.044	0.042	0.040	0.038	0.037	0.036	0.034	0.033	0.032	0.031	0.030	0.028
		out	0.505	0.127	0.100	0.070	0.062	0.059	0.051	0.046	0.043	0.041	0.039	0.037	0.035	0.035	0.033	0.031	0.031	0.028	0.029	0.029	0.029	0.029
	SN	in	0.000	0.061	0.066	0.065	0.059	0.055	0.052	0.047	0.044	0.041	0.039	0.036	0.034	0.032	0.031	0.030	0.029	0.028	0.027	0.025	0.024	0.023
		out	0.498	0.086	0.059	0.050	0.043	0.038	0.032	0.030	0.026	0.024	0.022	0.021	0.021	0.020	0.021	0.020	0.020	0.021	0.021	0.021	0.021	0.022
2-types	EC, IA	in	0.492	0.140	0.166	0.137	0.121	0.110	0.100	0.091	0.086	0.080	0.075	0.071	0.067	0.062	0.060	0.058	0.055	0.053	0.051	0.049	0.046	0.044
		out	0.481	0.082	0.090	0.046	0.042	0.034	0.031	0.030	0.028	0.028	0.029	0.030	0.031	0.032	0.033	0.035	0.037	0.039	0.039	0.041	0.043	0.045
	EC, RD	in	0.986	0.189	0.178	0.141	0.114	0.099	0.093	0.084	0.077	0.072	0.067	0.062	0.059	0.056	0.053	0.052	0.050	0.048	0.045	0.043	0.041	0.040
		out	0.377	0.044	0.111	0.064	0.027	0.027	0.028	0.029	0.031	0.032	0.033	0.034	0.035	0.036	0.038	0.040	0.041	0.043	0.044	0.046	0.048	0.050
	EC, LC	in	0.000	0.050	0.068	0.066	0.063	0.058	0.055	0.052	0.050	0.047	0.044	0.041	0.040	0.037	0.036	0.037	0.035	0.035	0.033	0.032	0.029	0.027
		out	0.505	0.127	0.092	0.055	0.045	0.041	0.037	0.034	0.032	0.032	0.031	0.030	0.029	0.029	0.028	0.028	0.027	0.027	0.027	0.029	0.028	0.028
	EC, SN	in	0.000	0.061	0.066	0.065	0.059	0.055	0.052	0.047	0.044	0.041	0.039	0.036	0.034	0.032	0.031	0.030	0.029	0.028	0.027	0.025	0.024	0.023
		out	0.498	0.086	0.059	0.050	0.043	0.038	0.032	0.030	0.026	0.024	0.022	0.021	0.021	0.020	0.021	0.020	0.020	0.021	0.021	0.021	0.021	0.022
	IA, RD	in	1.956	0.203	0.167	0.138	0.120	0.109	0.099	0.090	0.085	0.079	0.075	0.071	0.066	0.061	0.059	0.057	0.054	0.052	0.050	0.048	0.046	0.044
		out	1.488	0.160	0.091	0.047	0.042	0.035	0.033	0.031	0.030	0.029	0.030	0.030	0.032	0.032	0.034	0.035	0.037	0.038	0.039	0.041	0.042	0.044
	IA, LC	in	0.000	0.050	0.067	0.074	0.076	0.061	0.058	0.054	0.051	0.047	0.045	0.043	0.041	0.038	0.037	0.037	0.036	0.035	0.034	0.032	0.031	0.029
		out	0.509	0.126	0.087	0.054	0.048	0.045	0.039	0.035	0.034	0.033	0.031	0.031	0.030	0.030	0.029	0.028	0.028	0.028	0.028	0.029	0.029	0.029
	IA, SN	in	0.000	0.061	0.065	0.065	0.057	0.053	0.050	0.046	0.043	0.041	0.039	0.036	0.034	0.031	0.031	0.029	0.028	0.027	0.026	0.025	0.023	0.023
		out	0.498	0.084	0.058	0.046	0.036	0.032	0.029	0.026	0.023	0.022	0.020	0.020	0.019	0.019	0.020	0.020	0.020	0.020	0.020	0.020	0.021	0.021
	RD, LC	in	0.000	0.050	0.068	0.065	0.059	0.054	0.053	0.050	0.048	0.045	0.043	0.041	0.039	0.036	0.036	0.036	0.034	0.034	0.032	0.032	0.029	0.027
		out	0.505	0.127	0.080	0.060	0.052	0.046	0.041	0.037	0.034	0.033	0.032	0.031	0.029	0.029	0.028	0.028	0.027	0.026	0.027	0.028	0.028	0.028
	RD, SN	in	0.000	0.061	0.066	0.065	0.059	0.055	0.052	0.047	0.044	0.041	0.039	0.036	0.034	0.032	0.032	0.030	0.029	0.028	0.027	0.025	0.024	0.023
		out	0.502	0.086	0.059	0.050	0.042	0.038	0.032	0.030	0.026	0.024	0.022	0.021	0.021	0.020	0.021	0.020	0.021	0.021	0.021	0.021	0.021	0.022
	LC, SN	in	0.000	0.049	0.053	0.062	0.056	0.053	0.050	0.048	0.045	0.042	0.040	0.037	0.035	0.033	0.032	0.031	0.030	0.028	0.027	0.026	0.025	0.024
		out	0.641	0.129	0.084	0.051	0.044	0.040	0.034	0.031	0.028	0.026	0.024	0.024	0.024	0.023	0.022	0.022	0.022	0.023	0.023	0.024	0.025	0.025
3-types	EC, IA, RD	in	0.413	0.135	0.166	0.137	0.120	0.109	0.099	0.091	0.085	0.079	0.075	0.071	0.066	0.061	0.059	0.057	0.054	0.052	0.050	0.048	0.046	0.044
		out	0.501	0.082	0.090	0.047	0.041	0.034	0.032	0.030	0.029	0.028	0.029	0.029	0.031	0.032	0.033	0.035	0.037	0.038	0.039	0.041	0.042	0.044
	EC, IA, LC	in	0.000	0.050	0.066	0.074	0.077	0.061	0.058	0.054	0.051	0.047	0.046	0.043	0.041	0.038	0.037	0.037	0.036	0.036	0.034	0.033	0.031	0.029
		out	0.509	0.126	0.085	0.053	0.047	0.044	0.039	0.034	0.033	0.033	0.031	0.030	0.029	0.029	0.028	0.028	0.027	0.027	0.028	0.028	0.029	0.029
	EC, IA, SN	in	0.000	0.061	0.065	0.065	0.057	0.053	0.050	0.046	0.043	0.041	0.039	0.036	0.034	0.031	0.031	0.029	0.028	0.027	0.026	0.025	0.023	0.023
		out	0.498	0.084	0.058	0.046	0.036	0.032	0.029	0.026	0.023	0.022	0.020	0.020	0.019	0.019	0.020	0.020	0.020	0.020	0.020	0.020	0.021	0.021
	EC, RD, LC	in	0.000	0.050	0.067	0.066	0.062	0.057	0.055	0.052	0.050	0.047	0.044	0.041	0.039	0.036	0.036	0.036	0.034	0.033	0.031	0.031	0.029	0.027
		out	0.505	0.127	0.079	0.056	0.046	0.041	0.038	0.035	0.032	0.032	0.031	0.030	0.029	0.029	0.028	0.028	0.027	0.026	0.026	0.028	0.028	0.028
	EC, RD, SN	in	0.000	0.061	0.066	0.065	0.059	0.055	0.052	0.047	0.044	0.041	0.039	0.036	0.034	0.032	0.032	0.030	0.029	0.028	0.027	0.025	0.024	0.023
		out	0.502	0.086	0.059	0.050	0.042	0.038	0.032	0.030	0.026	0.024	0.022	0.021	0.021	0.020	0.021	0.020	0.021	0.021	0.021	0.021	0.021	0.022
	EC, LC, SN	in	0.000	0.049	0.053	0.062	0.056	0.053	0.050	0.048	0.045	0.042	0.040	0.037	0.035	0.033	0.032	0.031	0.030	0.028	0.027	0.026	0.025	0.024
		out	0.641	0.129	0.084	0.051	0.044	0.040	0.034	0.031	0.028	0.026	0.024	0.024	0.024	0.023	0.022	0.022	0.022	0.023	0.023	0.024	0.025	0.025
	IA, RD, LC	in	0.000	0.050	0.066	0.074	0.077	0.061	0.058	0.054	0.051	0.047	0.046	0.043	0.041	0.038	0.037	0.037	0.036	0.036	0.034	0.033	0.031	0.029
		out	0.509	0.126	0.070	0.052	0.047	0.043	0.039	0.035	0.034	0.033	0.031	0.030	0.029	0.029	0.028	0.028	0.027	0.027	0.028	0.028	0.029	0.029
	IA, RD, SN	in	0.000	0.061	0.065	0.065	0.057	0.053	0.050	0.046	0.043	0.041	0.039	0.036	0.034	0.031	0.031	0.029	0.028	0.027	0.026	0.025	0.023	0.022
		out	0.502	0.084	0.058	0.046	0.036	0.032	0.029	0.026	0.023	0.022	0.020	0.020	0.019	0.019	0.020	0.020	0.020	0.020	0.020	0.020	0.021	0.021
	IA, LC, SN	in	0.000	0.049	0.053	0.062	0.055	0.052	0.049	0.047	0.044	0.042	0.040	0.038	0.035	0.033	0.032	0.031	0.029	0.028	0.027	0.025	0.024	0.023
		out	0.641	0.128	0.084	0.048	0.038	0.034	0.031	0.029	0.026	0.025	0.024	0.023	0.023	0.023	0.022	0.022	0.022	0.023	0.023	0.023	0.024	0.024
	RD, LC, SN	in	0.000	0.049	0.053	0.062	0.056	0.053	0.050	0.048	0.045	0.042	0.040	0.037	0.035	0.033	0.032	0.031	0.030	0.028	0.027	0.026	0.025	0.024
		out	0.641	0.129	0.068	0.051	0.044	0.040	0.034	0.031	0.028	0.026	0.024	0.024	0.024	0.023	0.022	0.022	0.022	0.023	0.023	0.024	0.025	0.025
4-types	EC, IA, RD, LC	in	0.000	0.050	0.066	0.074	0.077	0.061	0.058	0.054	0.051	0.047	0.046	0.043	0.041	0.038	0.037	0.037	0.036	0.036	0.034	0.033	0.031	0.029
		out	0.509	0.126	0.070	0.051	0.047	0.043	0.039	0.034	0.033	0.033	0.031	0.030	0.029	0.029	0.028	0.028	0.027	0.027	0.028	0.028	0.029	0.029
	EC, IA, RD, SN	in	0.000	0.061	0.065	0.065	0.057	0.053	0.050	0.046	0.043	0.041	0.039	0.036	0.034	0.031	0.031	0.029	0.028	0.027	0.026	0.025	0.023	0.022
		out	0.502	0.084	0.058	0.046	0.036	0.032	0.029	0.026	0.023	0.022	0.020	0.020	0.019	0.019	0.020	0.020	0.020	0.020	0.020	0.020	0.021	0.021
	IA, RD, LC, SN	in	0.000	0.049	0.053	0.062	0.055	0.052	0.049	0.047	0.044	0.042	0.040	0.038	0.035	0.033	0.032	0.031	0.029	0.028	0.027	0.025	0.024	0.023
		out	0.641	0.128	0.068	0.048	0.038	0.034	0.031	0.029	0.026	0.025	0.024	0.023	0.023	0.023	0.022	0.022	0.022	0.023	0.023	0.023	0.024	0.024
	EC, RD, LC, SN	in	0.000	0.049	0.053	0.062	0.056	0.053	0.050	0.048	0.045	0.042	0.040	0.037	0.035	0.033	0.032	0.031	0.030	0.028	0.027	0.026	0.025	0.024
		out	0.641	0.129	0.068	0.051	0.044	0.040	0.034	0.031	0.028	0.026	0.024	0.024	0.024	0.023	0.022	0.022	0.022	0.023	0.023	0.024	0.025	0.025
	EC, IA, LC, SN	in	0.000	0.049	0.053	0.062	0.055	0.052	0.049	0.047	0.044	0.042	0.040	0.038	0.035	0.033	0.032	0.031	0.029	0.028	0.027	0.025	0.024	0.023
		out	0.641	0.128	0.084	0.048	0.038	0.034	0.031	0.029	0.026	0.025	0.024	0.023	0.023	0.023	0.022	0.022	0.022	0.023	0.023	0.023	0.024	0.024
5-types	EC, IA, RD, LC, SN	in	0.000	0.049	0.053	0.062	0.055	0.052	0.049	0.047	0.044	0.042	0.040	0.038	0.035	0.033	0.032	0.031	0.029	0.028	0.027	0.025	0.024	0.023
		out	0.641	0.128	0.068	0.048	0.038	0.034	0.031	0.029	0.026	0.025	0.024	0.023	0.023	0.023	0.022	0.022	0.022	0.023	0.023	0.023	0.024	0.024

Table 2.D.3: In- and out-of-sample accuracy standard errors - Game A (continued)

			Decisions used to estimate and classify																					
			(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)	(31)	(32)	(33)	(34)	(35)	(36)	(37)	(38)	(39)	(40)	(41)	(42)	(43)	(44)
1-type	EC	in	0.052	0.050	0.048	0.046	0.044	0.042	0.040	0.038	0.036	0.034	0.032	0.030	0.029	0.027	0.025	0.023	0.021	0.019	0.017	0.015	0.012	0.039
		out	0.054	0.057	0.060	0.063	0.066	0.069	0.072	0.076	0.080	0.083	0.089	0.094	0.101	0.108	0.116	0.126	0.137	0.155	0.177	0.207	0.256	1.725
	IA	in	0.042	0.041	0.040	0.038	0.036	0.034	0.032	0.031	0.029	0.027	0.026	0.025	0.023	0.022	0.021	0.019	0.018	0.016	0.015	0.013	0.011	0.035
		out	0.046	0.048	0.050	0.052	0.054	0.057	0.058	0.061	0.064	0.066	0.070	0.074	0.080	0.084	0.091	0.098	0.106	0.120	0.138	0.166	0.210	1.378
	RD	in	0.038	0.036	0.035	0.033	0.032	0.030	0.029	0.028	0.026	0.024	0.023	0.022	0.020	0.020	0.018	0.017	0.016	0.015	0.012	0.012	0.011	0.030
		out	0.051	0.054	0.056	0.059	0.060	0.063	0.064	0.067	0.071	0.072	0.077	0.080	0.085	0.089	0.096	0.102	0.110	0.123	0.139	0.162	0.201	1.319
	LC	in	0.027	0.026	0.025	0.024	0.023	0.022	0.021	0.020	0.019	0.018	0.018	0.017	0.016	0.015	0.014	0.013	0.012	0.011	0.011	0.010	0.008	0.038
		out	0.030	0.030	0.030	0.031	0.033	0.034	0.035	0.036	0.037	0.039	0.041	0.043	0.046	0.049	0.053	0.056	0.062	0.068	0.078	0.092	0.119	0.756
	SN	in	0.022	0.021	0.020	0.019	0.019	0.018	0.017	0.016	0.015	0.014	0.014	0.013	0.012	0.011	0.011	0.010	0.009	0.008	0.008	0.007	0.006	0.019
		out	0.023	0.023	0.024	0.025	0.026	0.027	0.028	0.029	0.031	0.033	0.035	0.037	0.039	0.041	0.045	0.048	0.052	0.057	0.067	0.078	0.102	0.662
2-types	EC, IA	in	0.042	0.041	0.040	0.038	0.036	0.034	0.032	0.031	0.029	0.027	0.026	0.025	0.023	0.022	0.021	0.019	0.018	0.016	0.015	0.013	0.011	0.035
		out	0.046	0.048	0.050	0.052	0.054	0.057	0.058	0.061	0.064	0.066	0.070	0.074	0.080	0.084	0.091	0.098	0.106	0.120	0.138	0.166	0.210	1.378
	EC, RD	in	0.038	0.036	0.035	0.033	0.032	0.030	0.029	0.028	0.026	0.024	0.023	0.022	0.020	0.020	0.018	0.017	0.016	0.015	0.012	0.012	0.011	0.030
		out	0.051	0.054	0.056	0.059	0.060	0.063	0.064	0.067	0.071	0.072	0.077	0.080	0.085	0.089	0.096	0.102	0.110	0.123	0.139	0.162	0.201	1.319
	EC, LC	in	0.027	0.026	0.024	0.022	0.021	0.020	0.019	0.019	0.018	0.016	0.016	0.015	0.014	0.013	0.012	0.011	0.010	0.009	0.008	0.007	0.006	0.022
		out	0.029	0.029	0.030	0.031	0.033	0.034	0.035	0.037	0.037	0.038	0.040	0.043	0.045	0.047	0.052	0.055	0.061	0.067	0.077	0.091	0.119	0.758
	EC, SN	in	0.022	0.021	0.020	0.019	0.019	0.018	0.017	0.016	0.015	0.014	0.014	0.013	0.012	0.011	0.011	0.010	0.009	0.008	0.008	0.007	0.006	0.019
		out	0.023	0.023	0.024	0.025	0.026	0.027	0.028	0.029	0.031	0.033	0.035	0.037	0.039	0.041	0.045	0.048	0.052	0.057	0.067	0.078	0.102	0.662
	IA, RD	in	0.042	0.040	0.039	0.038	0.036	0.034	0.032	0.031	0.029	0.027	0.026	0.024	0.023	0.022	0.021	0.019	0.017	0.016	0.015	0.013	0.011	0.035
		out	0.045	0.048	0.050	0.052	0.054	0.056	0.058	0.061	0.064	0.066	0.071	0.074	0.080	0.085	0.092	0.099	0.107	0.121	0.139	0.167	0.210	1.378
	IA, LC	in	0.028	0.027	0.024	0.023	0.022	0.021	0.020	0.019	0.018	0.017	0.016	0.015	0.014	0.013	0.012	0.011	0.010	0.010	0.009	0.008	0.007	0.022
		out	0.031	0.031	0.031	0.032	0.034	0.034	0.036	0.038	0.039	0.039	0.042	0.044	0.046	0.049	0.053	0.057	0.062	0.069	0.079	0.093	0.122	0.775
	IA, SN	in	0.021	0.021	0.020	0.019	0.018	0.017	0.016	0.016	0.015	0.014	0.013	0.013	0.012	0.011	0.010	0.009	0.009	0.008	0.007	0.006	0.006	0.019
		out	0.022	0.023	0.023	0.024	0.025	0.026	0.027	0.029	0.030	0.032	0.034	0.036	0.038	0.040	0.043	0.046	0.051	0.056	0.065	0.075	0.099	0.626
	RD, LC	in	0.026	0.025	0.024	0.022	0.021	0.020	0.019	0.019	0.018	0.016	0.016	0.015	0.014	0.013	0.012	0.011	0.010	0.009	0.008	0.007	0.006	0.023
		out	0.029	0.029	0.030	0.031	0.033	0.034	0.035	0.037	0.038	0.038	0.040	0.043	0.045	0.047	0.052	0.055	0.061	0.067	0.077	0.091	0.120	0.762
	RD, SN	in	0.022	0.021	0.020	0.019	0.019	0.018	0.017	0.016	0.015	0.014	0.014	0.013	0.012	0.011	0.011	0.010	0.009	0.008	0.008	0.007	0.006	0.019
		out	0.023	0.023	0.024	0.025	0.026	0.027	0.028	0.029	0.030	0.033	0.034	0.036	0.039	0.041	0.045	0.047	0.052	0.057	0.067	0.078	0.100	0.662
	LC, SN	in	0.023	0.022	0.021	0.020	0.019	0.018	0.017	0.016	0.015	0.015	0.014	0.013	0.013	0.012	0.011	0.010	0.009	0.008	0.008	0.007	0.006	0.022
		out	0.026	0.026	0.027	0.028	0.029	0.030	0.032	0.033	0.033	0.036	0.038	0.040	0.043	0.045	0.049	0.052	0.057	0.063	0.072	0.085	0.109	0.710
3-types	EC, IA, RD	in	0.042	0.040	0.039	0.038	0.036	0.034	0.032	0.031	0.029	0.027	0.026	0.024	0.023	0.022	0.021	0.019	0.017	0.016	0.015	0.013	0.011	0.035
		out	0.045	0.048	0.050	0.052	0.054	0.056	0.058	0.061	0.064	0.066	0.071	0.074	0.080	0.085	0.092	0.099	0.107	0.121	0.139	0.167	0.210	1.378
	EC, IA, LC	in	0.028	0.027	0.024	0.023	0.022	0.021	0.020	0.019	0.018	0.017	0.016	0.015	0.014	0.013	0.012	0.011	0.010	0.010	0.009	0.008	0.007	0.022
		out	0.031	0.031	0.031	0.032	0.034	0.034	0.036	0.038	0.039	0.039	0.041	0.044	0.046	0.049	0.053	0.057	0.062	0.069	0.079	0.093	0.122	0.775
	EC, IA, SN	in	0.021	0.021	0.020	0.019	0.018	0.017	0.016	0.016	0.015	0.014	0.013	0.013	0.012	0.011	0.010	0.009	0.009	0.008	0.007	0.006	0.006	0.019
		out	0.022	0.023	0.023	0.024	0.025	0.026	0.027	0.029	0.030	0.032	0.034	0.036	0.038	0.040	0.043	0.046	0.051	0.056	0.065	0.075	0.099	0.626
	EC, RD, LC	in	0.026	0.025	0.024	0.022	0.021	0.020	0.019	0.018	0.018	0.016	0.016	0.015	0.014	0.013	0.012	0.011	0.010	0.010	0.009	0.008	0.007	0.024
		out	0.028	0.029	0.030	0.031	0.033	0.034	0.035	0.036	0.037	0.039	0.041	0.043	0.045	0.047	0.052	0.056	0.062	0.068	0.077	0.092	0.117	0.744
	EC, RD, SN	in	0.022	0.021	0.020	0.019	0.019	0.018	0.017	0.016	0.015	0.014	0.014	0.013	0.012	0.011	0.011	0.010	0.009	0.008	0.008	0.007	0.006	0.019
		out	0.023	0.023	0.024	0.025	0.026	0.027	0.028	0.029	0.030	0.033	0.034	0.036	0.039	0.041	0.045	0.047	0.052	0.057	0.067	0.078	0.100	0.662
	EC, LC, SN	in	0.023	0.022	0.021	0.020	0.019	0.018	0.017	0.016	0.015	0.015	0.014	0.013	0.013	0.012	0.011	0.010	0.009	0.008	0.008	0.007	0.006	0.022
		out	0.026	0.026	0.027	0.028	0.029	0.030	0.032	0.033	0.033	0.036	0.038	0.040	0.043	0.045	0.049	0.052	0.057	0.063	0.072	0.085	0.109	0.710
	IA, RD, LC	in	0.028	0.027	0.024	0.023	0.022	0.021	0.020	0.019	0.018	0.017	0.016	0.015	0.014	0.013	0.012	0.011	0.010	0.010	0.009	0.008	0.007	0.022
		out	0.031	0.031	0.031	0.032	0.034	0.034	0.036	0.038	0.039	0.039	0.041	0.044	0.046	0.049	0.053	0.057	0.062	0.069	0.079	0.093	0.122	0.775
	IA, RD, SN	in	0.021	0.021	0.019	0.019	0.018	0.017	0.016	0.015	0.014	0.014	0.013	0.013	0.012	0.011	0.010	0.009	0.009	0.008	0.007	0.006	0.006	0.019
		out	0.022	0.023	0.023	0.024	0.025	0.026	0.027	0.029	0.030	0.032	0.034	0.036	0.038	0.040	0.044	0.047	0.051	0.056	0.066	0.076	0.099	0.626
	IA, LC, SN	in	0.022	0.021	0.020	0.019	0.018	0.017	0.016	0.016	0.015	0.014	0.013	0.012	0.012	0.011	0.010	0.009	0.009	0.008	0.007	0.007	0.006	0.022
		out	0.026	0.026	0.027	0.027	0.029	0.030	0.031	0.033	0.034	0.036	0.038	0.040	0.042	0.044	0.048	0.051	0.056	0.062	0.071	0.084	0.109	0.691
	RD, LC, SN	in	0.023	0.022	0.021	0.020	0.019	0.018	0.017	0.016	0.015	0.015	0.014	0.013	0.013	0.012	0.011	0.010	0.009	0.008	0.008	0.007	0.006	0.022
		out	0.026	0.026	0.027	0.028	0.029	0.030	0.032	0.033	0.033	0.036	0.038	0.040	0.043	0.045	0.049	0.052	0.057	0.063	0.072	0.085	0.109	0.710
4-types	EC, IA, RD, LC	in	0.028	0.027	0.024	0.023	0.022	0.021	0.020	0.019	0.018	0.017	0.016	0.015	0.014	0.013	0.012	0.011	0.010	0.010	0.009	0.008	0.007	0.022
		out	0.031	0.031	0.031	0.032	0.034	0.034	0.036	0.038	0.039	0.039	0.041	0.044	0.046	0.049	0.053	0.057	0.062	0.069	0.079	0.093	0.122	0.775
	EC, IA, RD, SN	in	0.021	0.021	0.019	0.019	0.018	0.017	0.016	0.015	0.014	0.014	0.013	0.013	0.012	0.011	0.010	0.009	0.009	0.008	0.007	0.006	0.006	0.019
		out	0.022	0.023	0.023	0.024	0.025	0.026	0.027	0.029	0.030	0.032	0.034	0.036	0.038	0.040	0.044	0.047	0.051	0.056	0.066	0.076	0.099	0.626
	IA, RD, LC, SN	in	0.022	0.021	0.020	0.019	0.018	0.017	0.016	0.016	0.015	0.014	0.013	0.012	0.012	0.011	0.010	0.009	0.009	0.008	0.007	0.007	0.006	0.022
		out	0.026	0.026	0.027	0.027	0.029	0.030	0.031	0.033	0.034	0.036	0.038	0.040	0.042	0.044	0.048	0.051	0.056	0.062	0.071	0.084	0.109	0.691
	EC, RD, LC, SN	in	0.023	0.022	0.021	0.020	0.019	0.018	0.017	0.016	0.015	0.015	0.014	0.013	0.013	0.012	0.011	0.010	0.009	0.008	0.008	0.007	0.006	0.022
		out	0.026	0.026	0.027	0.028	0.029	0.030	0.032	0.033	0.033	0.036	0.038	0.040	0.043	0.045	0.049	0.052	0.057	0.063	0.072	0.085	0.109	0.710
	EC, IA, LC, SN	in	0.022	0.021	0.020	0.019	0.018	0.017	0.016	0.016	0.015	0.014	0.013	0.012	0.012	0.011	0.010	0.009	0.009	0.008	0.007	0.007	0.006	0.022
		out	0.026	0.026	0.027	0.027	0.029	0.030	0.031	0.033	0.034	0.036	0.038	0.040	0.042	0.044	0.048	0.051	0.056	0.062	0.071	0.084	0.109	0.691
5-types	EC, IA, RD, LC, SN	in	0.022	0.021	0.020	0.019	0.018	0.017	0.016	0.016	0.015	0.014	0.013	0.012	0.012	0.011	0.010	0.009	0.009	0.008	0.007	0.007	0.006	0.022
		out	0.026	0.026	0.027	0.027	0.029	0.030	0.031	0.033	0.034	0.036	0.038	0.040	0.042	0.044	0.048	0.051	0.056	0.062	0.071	0.084	0.109	0.691

This table presents the standard errors of the in-sample and out-of-sample accuracies of all 31 worlds in Game A. Standard errors are derived as $se = \sigma_{\mu_{\text{acc}}}/\sqrt{S}$, where $\sigma_{\mu_{\text{acc}}}$ refers to the standard deviation of the mean accuracy μ_{acc} which equals the average of the $M = 1000$ computed accuracies. The accuracy itself is derived as the ratio of correctly fitted decisions over the total number of decisions. We assume Senders' first-order beliefs to equal one. EC reflects the standard economic model, IA inequity aversion, RD reference-dependent preferences, LC lying costs, and SN reputation or social norms. Each column corresponds to a specific number of decisions used to estimate individual parameters, classify participants into types and to compute the in-sample performance. The remaining decisions are used to derive the out-of-sample fit. All standard errors are multiplied by 100.

Table 2.D.4: In- and out-of-sample accuracy - Game B

			Decisions used to estimate and classify																					
			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)
1-type	EC	in	0.634	0.634	0.637	0.636	0.636	0.636	0.636	0.635	0.636	0.636	0.637	0.637	0.637	0.637	0.637	0.637	0.637	0.637	0.637	0.637	0.637	0.637
		out	0.637	0.637	0.637	0.637	0.638	0.637	0.638	0.638	0.638	0.638	0.638	0.637	0.637	0.637	0.637	0.637	0.637	0.637	0.637	0.637	0.637	0.637
	IA	in	0.955	0.910	0.888	0.864	0.852	0.841	0.834	0.828	0.825	0.821	0.819	0.817	0.815	0.814	0.813	0.812	0.811	0.810	0.809	0.808	0.808	0.807
		out	0.686	0.723	0.738	0.747	0.753	0.758	0.762	0.765	0.768	0.770	0.772	0.773	0.774	0.776	0.777	0.778	0.779	0.779	0.780	0.780	0.781	0.781
	RD	in	0.924	0.842	0.824	0.816	0.808	0.803	0.798	0.796	0.794	0.793	0.792	0.792	0.791	0.790	0.790	0.789	0.789	0.789	0.788	0.788	0.788	0.787
		out	0.678	0.723	0.742	0.750	0.756	0.760	0.763	0.766	0.768	0.769	0.771	0.772	0.774	0.774	0.775	0.776	0.777	0.777	0.778	0.778	0.779	0.779
	LC	in	1.000	0.996	0.988	0.967	0.947	0.931	0.917	0.906	0.897	0.889	0.883	0.878	0.874	0.870	0.867	0.864	0.862	0.860	0.858	0.857	0.855	0.854
		out	0.704	0.731	0.741	0.748	0.752	0.756	0.761	0.765	0.769	0.773	0.777	0.781	0.784	0.787	0.790	0.792	0.795	0.797	0.799	0.801	0.803	0.804
	SN	in	1.000	0.957	0.937	0.913	0.901	0.891	0.884	0.878	0.874	0.870	0.868	0.865	0.864	0.862	0.860	0.859	0.858	0.857	0.856	0.855	0.854	0.854
		out	0.715	0.745	0.768	0.780	0.790	0.796	0.802	0.806	0.810	0.812	0.814	0.817	0.818	0.820	0.821	0.822	0.823	0.824	0.825	0.826	0.827	0.828
2-types	EC, IA	in	0.729	0.725	0.852	0.846	0.839	0.832	0.827	0.826	0.823	0.820	0.818	0.816	0.815	0.814	0.813	0.812	0.811	0.810	0.809	0.808	0.808	0.807
		out	0.598	0.619	0.751	0.760	0.765	0.768	0.771	0.771	0.774	0.775	0.777	0.778	0.778	0.779	0.780	0.781	0.780	0.780	0.781	0.781	0.781	0.782
	EC, RD	in	0.697	0.655	0.744	0.770	0.774	0.784	0.783	0.786	0.788	0.788	0.791	0.791	0.790	0.790	0.789	0.789	0.789	0.789	0.789	0.788	0.788	0.787
		out	0.609	0.626	0.730	0.756	0.764	0.772	0.774	0.777	0.777	0.777	0.779	0.779	0.779	0.779	0.780	0.780	0.779	0.779	0.779	0.779	0.779	0.780
	EC, LC	in	1.000	0.996	0.986	0.949	0.902	0.895	0.884	0.879	0.874	0.874	0.867	0.866	0.860	0.858	0.856	0.854	0.852	0.851	0.847	0.850	0.849	0.848
		out	0.704	0.731	0.743	0.762	0.782	0.787	0.792	0.794	0.794	0.794	0.794	0.795	0.797	0.796	0.799	0.800	0.801	0.803	0.802	0.806	0.805	0.808
	EC, SN	in	1.000	0.957	0.937	0.913	0.901	0.891	0.884	0.878	0.874	0.870	0.868	0.865	0.864	0.862	0.860	0.859	0.858	0.857	0.856	0.855	0.854	0.854
		out	0.715	0.745	0.768	0.780	0.790	0.796	0.802	0.806	0.810	0.812	0.814	0.817	0.818	0.820	0.821	0.822	0.823	0.824	0.825	0.826	0.827	0.828
	IA, RD	in	0.956	0.910	0.887	0.864	0.852	0.841	0.835	0.829	0.826	0.823	0.821	0.819	0.818	0.817	0.816	0.814	0.813	0.813	0.812	0.811	0.811	0.810
		out	0.684	0.722	0.738	0.747	0.754	0.759	0.765	0.768	0.771	0.773	0.775	0.777	0.779	0.780	0.781	0.782	0.783	0.784	0.785	0.785	0.786	0.786
	IA, LC	in	1.000	0.996	0.982	0.940	0.915	0.900	0.890	0.883	0.875	0.869	0.865	0.862	0.859	0.857	0.854	0.853	0.850	0.850	0.848	0.849	0.847	0.846
		out	0.704	0.731	0.740	0.755	0.761	0.770	0.775	0.781	0.784	0.787	0.790	0.792	0.795	0.798	0.799	0.802	0.803	0.804	0.806	0.809	0.808	0.810
	IA, SN	in	1.000	0.957	0.936	0.912	0.898	0.889	0.881	0.874	0.870	0.866	0.864	0.861	0.860	0.858	0.857	0.856	0.854	0.853	0.852	0.852	0.851	0.850
		out	0.714	0.745	0.768	0.782	0.792	0.798	0.803	0.808	0.812	0.815	0.817	0.819	0.821	0.822	0.824	0.825	0.826	0.827	0.827	0.828	0.829	0.830
	RD, LC	in	1.000	0.996	0.986	0.963	0.932	0.919	0.906	0.897	0.886	0.884	0.874	0.871	0.865	0.860	0.857	0.855	0.853	0.851	0.847	0.850	0.849	0.848
		out	0.704	0.731	0.743	0.751	0.757	0.763	0.768	0.772	0.776	0.780	0.782	0.785	0.786	0.791	0.794	0.796	0.798	0.801	0.800	0.804	0.804	0.807
	RD, SN	in	1.000	0.957	0.937	0.913	0.901	0.891	0.884	0.878	0.874	0.870	0.868	0.865	0.863	0.862	0.860	0.859	0.857	0.856	0.855	0.855	0.854	0.853
		out	0.714	0.745	0.768	0.780	0.790	0.796	0.802	0.807	0.810	0.812	0.815	0.817	0.819	0.820	0.821	0.823	0.824	0.825	0.826	0.827	0.828	0.828
	LC, SN	in	1.000	0.996	0.979	0.939	0.919	0.906	0.898	0.891	0.886	0.882	0.880	0.876	0.872	0.871	0.868	0.866	0.865	0.863	0.862	0.862	0.861	0.860
		out	0.700	0.731	0.754	0.773	0.783	0.789	0.795	0.798	0.801	0.803	0.804	0.806	0.809	0.810	0.813	0.814	0.814	0.816	0.817	0.819	0.819	0.821
3-types	EC, IA, RD	in	0.735	0.725	0.852	0.846	0.840	0.833	0.829	0.827	0.824	0.822	0.820	0.818	0.818	0.816	0.815	0.814	0.813	0.813	0.812	0.811	0.811	0.810
		out	0.597	0.619	0.751	0.759	0.766	0.769	0.774	0.774	0.776	0.778	0.780	0.781	0.782	0.784	0.784	0.785	0.784	0.785	0.785	0.786	0.786	0.787
	EC, IA, LC	in	1.000	0.996	0.980	0.930	0.897	0.888	0.879	0.877	0.871	0.867	0.864	0.860	0.858	0.857	0.853	0.853	0.850	0.849	0.848	0.849	0.847	0.846
		out	0.704	0.731	0.742	0.764	0.775	0.782	0.788	0.790	0.791	0.792	0.795	0.797	0.799	0.800	0.800	0.803	0.804	0.805	0.806	0.809	0.808	0.810
	EC, IA, SN	in	1.000	0.957	0.936	0.912	0.898	0.889	0.881	0.874	0.870	0.866	0.864	0.861	0.860	0.858	0.857	0.856	0.854	0.853	0.852	0.852	0.851	0.850
		out	0.714	0.745	0.768	0.782	0.792	0.798	0.803	0.808	0.812	0.815	0.817	0.819	0.821	0.822	0.824	0.825	0.826	0.827	0.827	0.828	0.829	0.830
	EC, RD, LC	in	1.000	0.996	0.985	0.950	0.903	0.895	0.884	0.879	0.874	0.874	0.868	0.866	0.860	0.858	0.856	0.854	0.852	0.853	0.855	0.854	0.854	0.852
		out	0.704	0.731	0.744	0.762	0.781	0.786	0.792	0.794	0.794	0.794	0.794	0.795	0.797	0.796	0.799	0.800	0.801	0.804	0.808	0.809	0.809	0.811
	EC, RD, SN	in	1.000	0.957	0.937	0.913	0.901	0.891	0.884	0.878	0.874	0.870	0.868	0.865	0.863	0.862	0.860	0.859	0.857	0.856	0.855	0.855	0.854	0.853
		out	0.714	0.745	0.768	0.780	0.790	0.796	0.802	0.807	0.810	0.812	0.815	0.817	0.819	0.820	0.821	0.823	0.824	0.825	0.826	0.827	0.828	0.828
	EC, LC, SN	in	1.000	0.996	0.979	0.939	0.919	0.906	0.898	0.891	0.886	0.882	0.880	0.876	0.872	0.871	0.868	0.866	0.865	0.863	0.862	0.862	0.861	0.860
		out	0.700	0.731	0.754	0.773	0.783	0.789	0.795	0.798	0.801	0.803	0.804	0.806	0.809	0.810	0.813	0.814	0.814	0.816	0.817	0.819	0.819	0.821
	IA, RD, LC	in	1.000	0.996	0.982	0.940	0.914	0.900	0.890	0.883	0.875	0.869	0.865	0.862	0.859	0.857	0.854	0.853	0.850	0.849	0.848	0.849	0.847	0.846
		out	0.704	0.731	0.741	0.755	0.762	0.770	0.775	0.781	0.784	0.787	0.790	0.792	0.795	0.798	0.799	0.802	0.803	0.805	0.806	0.809	0.808	0.810
	IA, RD, SN	in	1.000	0.957	0.936	0.912	0.898	0.889	0.881	0.874	0.870	0.866	0.864	0.861	0.860	0.858	0.857	0.856	0.854	0.853	0.852	0.851	0.851	0.850
		out	0.713	0.745	0.768	0.782	0.792	0.798	0.803	0.808	0.812	0.815	0.818	0.819	0.821	0.822	0.824	0.825	0.826	0.827	0.828	0.829	0.830	0.831
	IA, LC, SN	in	1.000	0.996	0.979	0.937	0.916	0.903	0.893	0.887	0.882	0.878	0.876	0.871	0.869	0.867	0.865	0.863	0.861	0.860	0.859	0.858	0.858	0.857
		out	0.700	0.731	0.754	0.774	0.785	0.791	0.796	0.801	0.803	0.806	0.807	0.810	0.812	0.813	0.815	0.817	0.818	0.819	0.820	0.821	0.822	0.824
	RD, LC, SN	in	1.000	0.996	0.979	0.939	0.919	0.906	0.898	0.891	0.886	0.882	0.880	0.876	0.872	0.871	0.868	0.866	0.865	0.863	0.862	0.862	0.861	0.860
		out	0.700	0.731	0.754	0.773	0.783	0.789	0.795	0.798	0.801	0.803	0.804	0.806	0.809	0.810	0.813	0.814	0.814	0.816	0.817	0.819	0.819	0.821
4-types	EC, IA, RD, LC	in	1.000	0.996	0.980	0.930	0.897	0.888	0.879	0.877	0.871	0.867	0.864	0.860	0.858	0.857	0.853	0.853	0.850	0.849	0.848	0.849	0.847	0.846
		out	0.704	0.731	0.742	0.765	0.775	0.782	0.788	0.790	0.791	0.792	0.795	0.797	0.799	0.800	0.800	0.803	0.804	0.805	0.806	0.809	0.808	0.810
	EC, IA, RD, SN	in	1.000	0.957	0.936	0.912	0.898	0.889	0.881	0.874	0.870	0.866	0.864	0.861	0.860	0.858	0.857	0.856	0.854	0.853	0.852	0.851	0.851	0.850
		out	0.713	0.745	0.768	0.782	0.792	0.798	0.803	0.808	0.812	0.815	0.818	0.819	0.821	0.822	0.824	0.825	0.826	0.827	0.828	0.829	0.830	0.831
	IA, RD, LC, SN	in	1.000	0.996	0.979	0.937	0.916	0.903	0.893	0.887	0.882	0.878	0.876	0.871	0.869	0.867	0.865	0.863	0.861	0.860	0.859	0.858	0.858	0.857
		out	0.700	0.731	0.754	0.774	0.785	0.791	0.796	0.801	0.803	0.806	0.807	0.810	0.812	0.813	0.815	0.817	0.818	0.819	0.820	0.821	0.822	0.824
	EC, RD, LC, SN	in	1.000	0.996	0.979	0.939	0.919	0.906	0.898	0.891	0.886	0.882	0.880	0.876	0.872	0.871	0.868	0.866	0.865	0.863	0.862	0.862	0.861	0.860
		out	0.700	0.731	0.754	0.773	0.783	0.789	0.795	0.798	0.801	0.803	0.804	0.806	0.809	0.810	0.813	0.814	0.814	0.816	0.817	0.819	0.819	0.821
	EC, IA, LC, SN	in	1.000	0.996	0.979	0.937	0.916	0.903	0.893	0.887	0.882	0.878	0.876	0.871	0.869	0.867	0.865	0.863	0.861	0.860	0.859	0.858	0.858	0.857
		out	0.700	0.731	0.754	0.774	0.785	0.791	0.796	0.801	0.803	0.806	0.807	0.810	0.812	0.813	0.815	0.817	0.818	0.819	0.820	0.821	0.822	0.824
5-types	EC, IA, RD, LC, SN	in	1.000	0.996	0.979	0.937	0.916	0.903	0.893	0.887	0.882	0.878	0.876	0.871	0.869	0.867	0.865	0.863	0.861	0.860	0.859	0.858	0.858	0.857
		out	0.700	0.731	0.754	0.774	0.785	0.791	0.796	0.801	0.803	0.806	0.807	0.810	0.812	0.813	0.815	0.817	0.818	0.819	0.820	0.821	0.822	0.824

Table 2.D.4: In- and out-of-sample accuracy - Game B (continued)

			Decisions used to estimate and classify																					
			(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)	(31)	(32)	(33)	(34)	(35)	(36)	(37)	(38)	(39)	(40)	(41)	(42)	(43)	(44)
1-type	EC	in	0.637	0.637	0.638	0.637	0.638	0.638	0.638	0.637	0.637	0.637	0.637	0.638	0.637	0.637	0.637	0.637	0.637	0.637	0.638	0.637	0.638	0.637
		out	0.638	0.637	0.637	0.637	0.637	0.637	0.637	0.637	0.637	0.637	0.637	0.637	0.637	0.637	0.637	0.638	0.638	0.636	0.635	0.635	0.633	0.632
	IA	in	0.806	0.806	0.805	0.805	0.804	0.804	0.804	0.803	0.803	0.802	0.802	0.802	0.801	0.801	0.800	0.800	0.800	0.800	0.799	0.799	0.799	0.798
		out	0.782	0.782	0.782	0.783	0.783	0.783	0.783	0.783	0.784	0.784	0.784	0.784	0.784	0.784	0.785	0.785	0.785	0.784	0.783	0.784	0.781	0.782
	RD	in	0.787	0.787	0.787	0.787	0.787	0.787	0.787	0.787	0.786	0.786	0.786	0.786	0.786	0.786	0.786	0.786	0.786	0.786	0.786	0.786	0.786	0.786
		out	0.780	0.780	0.780	0.781	0.781	0.781	0.781	0.782	0.782	0.782	0.782	0.782	0.783	0.783	0.783	0.784	0.784	0.783	0.782	0.783	0.781	0.781
	LC	in	0.853	0.852	0.851	0.851	0.850	0.849	0.848	0.848	0.847	0.846	0.846	0.846	0.845	0.845	0.844	0.844	0.844	0.844	0.843	0.843	0.843	0.843
		out	0.806	0.808	0.809	0.809	0.810	0.811	0.812	0.813	0.814	0.815	0.816	0.816	0.816	0.817	0.818	0.819	0.819	0.820	0.820	0.821	0.821	0.823
	SN	in	0.853	0.853	0.852	0.852	0.851	0.851	0.850	0.850	0.850	0.849	0.849	0.849	0.849	0.849	0.848	0.848	0.848	0.848	0.848	0.848	0.848	0.848
		out	0.828	0.829	0.829	0.830	0.830	0.830	0.831	0.832	0.832	0.833	0.833	0.833	0.833	0.834	0.834	0.835	0.835	0.835	0.835	0.835	0.834	0.834
2-types	EC, IA	in	0.806	0.806	0.805	0.805	0.805	0.804	0.804	0.803	0.803	0.802	0.802	0.802	0.801	0.801	0.800	0.800	0.800	0.800	0.799	0.799	0.799	0.798
		out	0.782	0.783	0.783	0.783	0.783	0.783	0.784	0.784	0.784	0.784	0.784	0.784	0.784	0.784	0.785	0.785	0.785	0.784	0.783	0.784	0.781	0.782
	EC, RD	in	0.787	0.787	0.787	0.787	0.787	0.787	0.787	0.787	0.786	0.786	0.786	0.786	0.786	0.786	0.786	0.786	0.786	0.786	0.786	0.786	0.786	0.786
		out	0.780	0.780	0.780	0.781	0.781	0.781	0.781	0.782	0.782	0.782	0.782	0.782	0.783	0.783	0.783	0.784	0.784	0.783	0.782	0.783	0.781	0.781
	EC, LC	in	0.847	0.847	0.846	0.846	0.847	0.845	0.844	0.846	0.845	0.843	0.844	0.844	0.843	0.843	0.843	0.843	0.843	0.842	0.842	0.842	0.842	0.842
		out	0.807	0.808	0.810	0.810	0.812	0.811	0.812	0.814	0.815	0.814	0.816	0.817	0.817	0.818	0.819	0.819	0.820	0.820	0.821	0.822	0.822	0.823
	EC, SN	in	0.853	0.853	0.852	0.852	0.851	0.851	0.850	0.850	0.850	0.849	0.849	0.849	0.849	0.849	0.848	0.848	0.848	0.848	0.848	0.848	0.848	0.848
		out	0.828	0.829	0.829	0.830	0.830	0.830	0.831	0.832	0.832	0.833	0.833	0.833	0.833	0.834	0.834	0.835	0.835	0.835	0.835	0.835	0.834	0.834
	IA, RD	in	0.809	0.809	0.809	0.808	0.808	0.808	0.807	0.807	0.807	0.806	0.806	0.806	0.806	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.804
		out	0.787	0.787	0.788	0.789	0.789	0.789	0.790	0.790	0.790	0.791	0.791	0.791	0.791	0.792	0.792	0.793	0.793	0.793	0.793	0.794	0.792	0.794
	IA, LC	in	0.847	0.846	0.845	0.845	0.844	0.844	0.844	0.845	0.844	0.842	0.844	0.843	0.843	0.843	0.842	0.842	0.842	0.842	0.842	0.841	0.841	0.841
		out	0.811	0.813	0.814	0.814	0.814	0.815	0.816	0.818	0.819	0.818	0.820	0.820	0.821	0.822	0.823	0.824	0.825	0.825	0.826	0.827	0.827	0.831
	IA, SN	in	0.850	0.850	0.849	0.849	0.849	0.848	0.848	0.847	0.847	0.847	0.847	0.846	0.846	0.846	0.846	0.846	0.846	0.846	0.846	0.846	0.846	0.846
		out	0.831	0.831	0.831	0.832	0.832	0.832	0.833	0.832	0.833	0.833	0.833	0.833	0.834	0.834	0.835	0.835	0.836	0.835	0.835	0.836	0.834	0.836
	RD, LC	in	0.847	0.847	0.846	0.846	0.847	0.845	0.844	0.846	0.845	0.842	0.844	0.844	0.843	0.843	0.843	0.843	0.842	0.842	0.842	0.842	0.842	0.842
		out	0.807	0.808	0.809	0.810	0.811	0.810	0.811	0.814	0.814	0.813	0.816	0.816	0.817	0.818	0.818	0.819	0.819	0.820	0.820	0.821	0.821	0.823
	RD, SN	in	0.853	0.852	0.852	0.851	0.851	0.851	0.850	0.850	0.850	0.849	0.849	0.849	0.849	0.848	0.848	0.848	0.848	0.848	0.848	0.848	0.848	0.848
		out	0.829	0.830	0.830	0.830	0.831	0.831	0.832	0.832	0.833	0.833	0.834	0.834	0.834	0.835	0.835	0.836	0.836	0.836	0.836	0.837	0.836	0.835
	LC, SN	in	0.860	0.859	0.859	0.858	0.858	0.857	0.857	0.856	0.856	0.855	0.855	0.855	0.855	0.854	0.854	0.854	0.854	0.854	0.853	0.853	0.853	0.853
		out	0.821	0.823	0.824	0.825	0.826	0.826	0.828	0.829	0.830	0.830	0.831	0.831	0.832	0.832	0.833	0.834	0.835	0.835	0.836	0.837	0.838	0.837
3-types	EC, IA, RD	in	0.809	0.809	0.809	0.808	0.808	0.808	0.807	0.807	0.807	0.807	0.806	0.806	0.806	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.804
		out	0.787	0.788	0.789	0.789	0.789	0.789	0.790	0.790	0.790	0.791	0.791	0.791	0.791	0.792	0.792	0.793	0.793	0.793	0.793	0.794	0.792	0.794
	EC, IA, LC	in	0.847	0.846	0.845	0.845	0.844	0.844	0.844	0.845	0.844	0.842	0.844	0.843	0.843	0.843	0.842	0.842	0.842	0.842	0.842	0.841	0.841	0.841
		out	0.812	0.813	0.814	0.814	0.814	0.815	0.816	0.818	0.819	0.818	0.820	0.820	0.821	0.822	0.823	0.824	0.825	0.825	0.826	0.827	0.827	0.831
	EC, IA, SN	in	0.850	0.850	0.849	0.849	0.849	0.848	0.848	0.847	0.847	0.847	0.847	0.846	0.846	0.846	0.846	0.846	0.846	0.846	0.846	0.846	0.846	0.846
		out	0.831	0.831	0.831	0.832	0.832	0.832	0.833	0.832	0.833	0.833	0.833	0.833	0.834	0.834	0.835	0.835	0.836	0.835	0.835	0.836	0.834	0.836
	EC, RD, LC	in	0.852	0.852	0.849	0.849	0.848	0.847	0.847	0.846	0.846	0.845	0.844	0.844	0.843	0.843	0.843	0.843	0.843	0.842	0.842	0.842	0.842	0.842
		out	0.810	0.812	0.811	0.811	0.811	0.812	0.812	0.813	0.814	0.816	0.816	0.816	0.817	0.818	0.818	0.819	0.819	0.819	0.820	0.821	0.822	0.823
	EC, RD, SN	in	0.853	0.852	0.852	0.851	0.851	0.851	0.850	0.850	0.850	0.849	0.849	0.849	0.849	0.848	0.848	0.848	0.848	0.848	0.848	0.848	0.848	0.848
		out	0.829	0.830	0.830	0.830	0.831	0.831	0.832	0.832	0.833	0.833	0.834	0.834	0.834	0.835	0.835	0.836	0.836	0.836	0.836	0.837	0.836	0.835
	EC, LC, SN	in	0.860	0.859	0.859	0.858	0.858	0.857	0.857	0.856	0.856	0.855	0.855	0.855	0.855	0.854	0.854	0.854	0.854	0.854	0.853	0.853	0.853	0.853
		out	0.821	0.823	0.824	0.825	0.826	0.826	0.828	0.829	0.830	0.830	0.831	0.831	0.832	0.832	0.833	0.834	0.835	0.835	0.836	0.837	0.838	0.837
	IA, RD, LC	in	0.847	0.846	0.845	0.845	0.844	0.844	0.844	0.845	0.844	0.842	0.844	0.843	0.843	0.843	0.842	0.842	0.842	0.842	0.842	0.841	0.841	0.841
		out	0.812	0.813	0.814	0.814	0.814	0.815	0.816	0.818	0.819	0.818	0.820	0.820	0.821	0.822	0.823	0.824	0.825	0.825	0.826	0.827	0.827	0.831
	IA, RD, SN	in	0.850	0.850	0.849	0.849	0.849	0.848	0.848	0.847	0.847	0.847	0.847	0.846	0.846	0.846	0.846	0.846	0.846	0.846	0.846	0.846	0.846	0.846
		out	0.831	0.831	0.832	0.832	0.832	0.833	0.833	0.833	0.833	0.834	0.834	0.834	0.834	0.835	0.836	0.836	0.837	0.837	0.837	0.837	0.837	0.836
	IA, LC, SN	in	0.856	0.856	0.855	0.855	0.855	0.854	0.854	0.853	0.853	0.852	0.852	0.852	0.851	0.851	0.852	0.852	0.852	0.851	0.851	0.851	0.851	0.851
		out	0.823	0.825	0.826	0.827	0.827	0.828	0.829	0.829	0.830	0.830	0.831	0.831	0.831	0.832	0.834	0.834	0.835	0.835	0.836	0.837	0.838	0.838
	RD, LC, SN	in	0.860	0.859	0.859	0.858	0.858	0.857	0.857	0.856	0.856	0.855	0.855	0.855	0.855	0.854	0.854	0.854	0.854	0.854	0.853	0.853	0.853	0.853
		out	0.821	0.823	0.824	0.825	0.826	0.826	0.828	0.829	0.830	0.830	0.831	0.831	0.832	0.832	0.833	0.834	0.835	0.835	0.836	0.837	0.838	0.837
4-types	EC, IA, RD, LC	in	0.847	0.846	0.845	0.845	0.844	0.844	0.844	0.845	0.844	0.842	0.844	0.843	0.843	0.843	0.842	0.842	0.842	0.842	0.842	0.841	0.841	0.841
		out	0.812	0.813	0.814	0.814	0.814	0.815	0.816	0.818	0.819	0.818	0.820	0.820	0.821	0.822	0.823	0.824	0.825	0.825	0.826	0.827	0.827	0.831
	EC, IA, RD, SN	in	0.850	0.850	0.849	0.849	0.849	0.848	0.848	0.847	0.847	0.847	0.847	0.846	0.846	0.846	0.846	0.846	0.846	0.846	0.846	0.846	0.846	0.846
		out	0.831	0.831	0.832	0.832	0.832	0.833	0.833	0.833	0.833	0.834	0.834	0.834	0.834	0.835	0.836	0.836	0.837	0.837	0.837	0.837	0.837	0.836
	IA, RD, LC, SN	in	0.856	0.856	0.855	0.855	0.855	0.854	0.854	0.853	0.853	0.852	0.852	0.852	0.851	0.851	0.852	0.852	0.852	0.851	0.851	0.851	0.851	0.851
		out	0.823	0.825	0.826	0.827	0.827	0.828	0.829	0.829	0.830	0.830	0.831	0.831	0.831	0.832	0.834	0.834	0.835	0.835	0.836	0.837	0.838	0.838
	EC, RD, LC, SN	in	0.860	0.859	0.859	0.858	0.858	0.857	0.857	0.856	0.856	0.855	0.855	0.855	0.855	0.854	0.854	0.854	0.854	0.854	0.853	0.853	0.853	0.853
		out	0.821	0.823	0.824	0.825	0.826	0.826	0.828	0.829	0.830	0.830	0.831	0.831	0.832	0.832	0.833	0.834	0.835	0.835	0.836	0.837	0.838	0.837
	EC, IA, LC, SN	in	0.856	0.856	0.855	0.855	0.855	0.854	0.854	0.853	0.853	0.852	0.852	0.852	0.851	0.851	0.852	0.852	0.852	0.851	0.851	0.851	0.851	0.851
		out	0.823	0.825	0.826	0.827	0.827	0.828	0.829	0.829	0.830	0.830	0.831	0.831	0.831	0.832	0.834	0.834	0.835	0.835	0.836	0.837	0.838	0.838
5-types	EC, IA, RD, LC, SN	in	0.856	0.856	0.855	0.855	0.855	0.854	0.854	0.853	0.853	0.852	0.852	0.852	0.851	0.851	0.852	0.852	0.852	0.851	0.851	0.851	0.851	0.851
		out	0.823	0.825	0.826	0.827	0.827	0.828	0.829	0.829	0.830	0.830	0.831	0.831	0.831	0.832	0.834	0.834	0.835	0.835	0.836	0.837	0.838	0.838

This table presents the accuracies of the in-sample and out-of-sample fits of all 31 worlds in Game B. The accuracy is derived as the ratio of correctly fitted decisions over the total number of decisions, whereas $M = 1000$ random selections were used. We assume Senders' first-order beliefs to equal one. EC reflects the standard economic model, IA inequity aversion, RD reference-dependent preferences, LC lying costs, and SN reputation or social norms. Each column corresponds to a specific number of decisions used to estimate individual parameters, classify participants into types and to compute the in-sample performance. The remaining decisions are used to derive the out-of-sample fit.

Table 2.D.5: In- and out-of-sample accuracy standard errors - Game B

			Decisions used to estimate and classify																					
			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)
1-type	EC	in	1.852	0.276	0.216	0.186	0.167	0.151	0.137	0.125	0.116	0.108	0.100	0.093	0.089	0.085	0.080	0.078	0.075	0.072	0.067	0.064	0.062	0.060
		out	0.042	0.013	0.015	0.018	0.021	0.023	0.025	0.027	0.029	0.031	0.032	0.034	0.036	0.038	0.040	0.043	0.045	0.048	0.049	0.052	0.054	0.057
	IA	in	1.464	0.185	0.172	0.152	0.135	0.125	0.116	0.106	0.099	0.094	0.087	0.082	0.079	0.073	0.071	0.069	0.065	0.063	0.060	0.058	0.056	0.053
		out	0.747	0.122	0.083	0.063	0.055	0.048	0.043	0.039	0.036	0.034	0.035	0.035	0.035	0.036	0.037	0.039	0.041	0.042	0.043	0.045	0.048	0.050
	RD	in	1.889	0.176	0.179	0.141	0.130	0.122	0.113	0.103	0.096	0.091	0.086	0.081	0.077	0.073	0.070	0.068	0.065	0.063	0.059	0.057	0.055	0.054
		out	0.764	0.118	0.083	0.055	0.040	0.035	0.031	0.029	0.029	0.029	0.030	0.031	0.032	0.033	0.035	0.037	0.039	0.041	0.042	0.045	0.048	0.051
	LC	in	0.000	0.068	0.087	0.079	0.071	0.064	0.060	0.055	0.053	0.050	0.047	0.043	0.042	0.039	0.038	0.036	0.035	0.033	0.032	0.030	0.030	0.029
		out	0.700	0.142	0.102	0.070	0.064	0.061	0.055	0.052	0.048	0.045	0.043	0.042	0.041	0.041	0.038	0.037	0.036	0.033	0.033	0.032	0.032	0.032
	SN	in	0.000	0.064	0.063	0.068	0.062	0.058	0.054	0.050	0.045	0.043	0.041	0.038	0.036	0.035	0.034	0.033	0.031	0.030	0.028	0.027	0.026	0.025
		out	0.478	0.089	0.073	0.064	0.054	0.050	0.043	0.036	0.034	0.031	0.029	0.027	0.027	0.027	0.027	0.027	0.026	0.026	0.026	0.026	0.027	0.028
2-types	EC, IA	in	0.592	0.152	0.160	0.149	0.133	0.125	0.116	0.106	0.098	0.094	0.087	0.082	0.079	0.073	0.071	0.069	0.065	0.063	0.060	0.058	0.056	0.053
		out	0.619	0.099	0.075	0.056	0.049	0.042	0.036	0.033	0.031	0.030	0.031	0.031	0.032	0.033	0.035	0.038	0.040	0.042	0.042	0.045	0.047	0.050
	EC, RD	in	0.906	0.203	0.188	0.162	0.149	0.130	0.121	0.109	0.101	0.093	0.087	0.082	0.077	0.073	0.071	0.069	0.065	0.063	0.059	0.057	0.055	0.054
		out	0.507	0.058	0.055	0.039	0.026	0.026	0.026	0.028	0.029	0.029	0.030	0.032	0.033	0.034	0.036	0.039	0.040	0.042	0.043	0.045	0.048	0.051
	EC, LC	in	0.000	0.068	0.087	0.070	0.070	0.065	0.061	0.056	0.053	0.050	0.048	0.044	0.043	0.041	0.039	0.038	0.037	0.035	0.034	0.032	0.030	0.029
		out	0.700	0.142	0.098	0.065	0.047	0.044	0.039	0.039	0.036	0.037	0.036	0.035	0.033	0.034	0.033	0.032	0.032	0.030	0.030	0.030	0.030	0.030
	EC, SN	in	0.000	0.064	0.063	0.068	0.062	0.058	0.054	0.050	0.045	0.043	0.041	0.038	0.036	0.035	0.034	0.033	0.031	0.030	0.028	0.027	0.026	0.025
		out	0.478	0.089	0.073	0.064	0.054	0.050	0.043	0.036	0.034	0.031	0.029	0.027	0.027	0.027	0.027	0.027	0.026	0.026	0.026	0.026	0.027	0.028
	IA, RD	in	1.430	0.184	0.172	0.152	0.134	0.124	0.113	0.104	0.097	0.093	0.086	0.081	0.077	0.072	0.070	0.068	0.064	0.062	0.059	0.057	0.055	0.053
		out	0.779	0.122	0.083	0.063	0.055	0.047	0.043	0.040	0.036	0.035	0.035	0.035	0.035	0.036	0.037	0.039	0.040	0.041	0.042	0.044	0.046	0.049
	IA, LC	in	0.000	0.068	0.084	0.085	0.085	0.074	0.067	0.061	0.059	0.057	0.054	0.049	0.048	0.044	0.044	0.041	0.040	0.037	0.036	0.033	0.032	0.031
		out	0.700	0.142	0.095	0.057	0.053	0.047	0.042	0.041	0.038	0.036	0.035	0.034	0.034	0.035	0.034	0.034	0.033	0.032	0.032	0.031	0.032	0.033
	IA, SN	in	0.000	0.064	0.063	0.065	0.060	0.056	0.053	0.050	0.046	0.044	0.042	0.040	0.038	0.035	0.034	0.033	0.031	0.031	0.029	0.028	0.026	0.025
		out	0.491	0.088	0.071	0.059	0.050	0.047	0.041	0.036	0.034	0.031	0.029	0.027	0.026	0.027	0.026	0.027	0.027	0.026	0.027	0.027	0.027	0.028
	RD, LC	in	0.000	0.068	0.086	0.077	0.069	0.062	0.059	0.054	0.052	0.049	0.047	0.044	0.043	0.040	0.039	0.038	0.036	0.035	0.034	0.032	0.030	0.029
		out	0.700	0.142	0.091	0.068	0.057	0.054	0.049	0.047	0.043	0.042	0.039	0.038	0.037	0.037	0.035	0.034	0.034	0.031	0.031	0.031	0.031	0.031
	RD, SN	in	0.000	0.064	0.063	0.068	0.062	0.058	0.054	0.050	0.045	0.043	0.041	0.038	0.036	0.035	0.034	0.033	0.031	0.030	0.028	0.027	0.026	0.025
		out	0.493	0.089	0.073	0.063	0.054	0.050	0.043	0.036	0.034	0.031	0.029	0.027	0.027	0.027	0.027	0.027	0.027	0.026	0.026	0.026	0.027	0.028
	LC, SN	in	0.000	0.067	0.066	0.062	0.060	0.056	0.053	0.049	0.046	0.043	0.041	0.038	0.037	0.036	0.035	0.033	0.031	0.030	0.029	0.027	0.026	0.025
		out	0.769	0.143	0.079	0.066	0.057	0.052	0.046	0.039	0.037	0.034	0.031	0.029	0.029	0.029	0.029	0.028	0.029	0.028	0.028	0.028	0.027	0.028
3-types	EC, IA, RD	in	0.557	0.152	0.160	0.149	0.132	0.124	0.114	0.104	0.097	0.092	0.085	0.081	0.077	0.072	0.069	0.067	0.064	0.062	0.059	0.057	0.055	0.052
		out	0.641	0.099	0.075	0.055	0.048	0.041	0.037	0.034	0.032	0.031	0.031	0.032	0.033	0.033	0.035	0.037	0.039	0.041	0.042	0.044	0.046	0.049
	EC, IA, LC	in	0.000	0.068	0.083	0.079	0.084	0.075	0.067	0.060	0.058	0.056	0.054	0.049	0.048	0.044	0.044	0.041	0.040	0.037	0.036	0.033	0.033	0.031
		out	0.700	0.142	0.095	0.055	0.050	0.043	0.038	0.038	0.035	0.034	0.033	0.033	0.032	0.034	0.033	0.033	0.033	0.032	0.031	0.031	0.032	0.033
	EC, IA, SN	in	0.000	0.064	0.063	0.065	0.060	0.056	0.053	0.050	0.046	0.044	0.042	0.040	0.038	0.035	0.034	0.033	0.031	0.031	0.029	0.028	0.026	0.025
		out	0.491	0.088	0.071	0.059	0.050	0.047	0.041	0.036	0.034	0.031	0.029	0.027	0.026	0.027	0.026	0.027	0.027	0.026	0.027	0.027	0.027	0.028
	EC, RD, LC	in	0.000	0.068	0.085	0.070	0.069	0.065	0.061	0.056	0.053	0.050	0.048	0.044	0.043	0.041	0.039	0.037	0.036	0.034	0.032	0.031	0.029	0.028
		out	0.700	0.142	0.088	0.064	0.048	0.044	0.040	0.039	0.036	0.037	0.036	0.035	0.033	0.034	0.033	0.032	0.032	0.030	0.029	0.029	0.029	0.029
	EC, RD, SN	in	0.000	0.064	0.063	0.068	0.062	0.058	0.054	0.050	0.045	0.043	0.041	0.038	0.036	0.035	0.034	0.033	0.031	0.030	0.028	0.027	0.026	0.025
		out	0.493	0.089	0.073	0.063	0.054	0.050	0.043	0.036	0.034	0.031	0.029	0.027	0.027	0.027	0.027	0.027	0.027	0.026	0.026	0.026	0.027	0.028
	EC, LC, SN	in	0.000	0.067	0.066	0.062	0.060	0.056	0.053	0.049	0.046	0.043	0.041	0.038	0.037	0.036	0.035	0.033	0.031	0.030	0.029	0.027	0.026	0.025
		out	0.769	0.143	0.079	0.066	0.057	0.052	0.046	0.039	0.037	0.034	0.031	0.029	0.029	0.029	0.029	0.028	0.029	0.028	0.028	0.028	0.027	0.028
	IA, RD, LC	in	0.000	0.068	0.084	0.085	0.085	0.074	0.067	0.061	0.059	0.057	0.054	0.049	0.048	0.044	0.044	0.041	0.040	0.037	0.036	0.033	0.033	0.031
		out	0.700	0.142	0.086	0.056	0.053	0.047	0.042	0.041	0.038	0.036	0.035	0.034	0.034	0.035	0.034	0.034	0.033	0.032	0.031	0.031	0.032	0.033
	IA, RD, SN	in	0.000	0.064	0.063	0.065	0.060	0.056	0.053	0.050	0.046	0.044	0.042	0.040	0.037	0.035	0.034	0.033	0.031	0.030	0.029	0.028	0.026	0.025
		out	0.505	0.088	0.071	0.059	0.050	0.047	0.041	0.036	0.034	0.031	0.028	0.027	0.026	0.026	0.026	0.027	0.027	0.026	0.026	0.027	0.027	0.028
	IA, LC, SN	in	0.000	0.067	0.066	0.061	0.058	0.054	0.052	0.049	0.046	0.044	0.042	0.040	0.038	0.036	0.035	0.033	0.031	0.031	0.029	0.028	0.027	0.025
		out	0.776	0.143	0.079	0.063	0.054	0.049	0.043	0.038	0.035	0.033	0.030	0.028	0.028	0.028	0.028	0.028	0.029	0.028	0.028	0.028	0.028	0.029
	RD, LC, SN	in	0.000	0.067	0.066	0.062	0.060	0.056	0.053	0.049	0.046	0.043	0.041	0.038	0.037	0.036	0.035	0.033	0.031	0.030	0.029	0.027	0.026	0.025
		out	0.769	0.143	0.079	0.065	0.057	0.052	0.046	0.039	0.037	0.034	0.031	0.029	0.029	0.029	0.029	0.028	0.029	0.028	0.028	0.028	0.027	0.028
4-types	EC, IA, RD, LC	in	0.000	0.068	0.083	0.079	0.084	0.075	0.067	0.060	0.058	0.056	0.054	0.049	0.048	0.044	0.044	0.041	0.040	0.037	0.036	0.033	0.033	0.031
		out	0.700	0.142	0.086	0.054	0.050	0.043	0.038	0.038	0.035	0.034	0.033	0.033	0.032	0.034	0.033	0.033	0.033	0.032	0.031	0.031	0.032	0.033
	EC, IA, RD, SN	in	0.000	0.064	0.063	0.065	0.060	0.056	0.053	0.050	0.046	0.044	0.042	0.040	0.037	0.035	0.034	0.033	0.031	0.030	0.029	0.028	0.026	0.025
		out	0.505	0.088	0.071	0.059	0.050	0.047	0.041	0.036	0.034	0.031	0.028	0.027	0.026	0.026	0.026	0.027	0.027	0.026	0.026	0.027	0.027	0.028
	IA, RD, LC, SN	in	0.000	0.067	0.066	0.061	0.058	0.054	0.052	0.049	0.046	0.044	0.042	0.040	0.038	0.036	0.035	0.033	0.031	0.031	0.029	0.028	0.027	0.025
		out	0.776	0.143	0.079	0.062	0.054	0.049	0.043	0.038	0.035	0.033	0.030	0.028	0.028	0.028	0.028	0.028	0.029	0.028	0.028	0.028	0.028	0.029
	EC, RD, LC, SN	in	0.000	0.067	0.066	0.062	0.060	0.056	0.053	0.049	0.046	0.043	0.041	0.038	0.037	0.036	0.035	0.033	0.031	0.030	0.029	0.027	0.026	0.025
		out	0.769	0.143	0.079	0.065	0.057	0.052	0.046	0.039	0.037	0.034	0.031	0.029	0.029	0.029	0.029	0.028	0.029	0.028	0.028	0.028	0.027	0.028
	EC, IA, LC, SN	in	0.000	0.067	0.066	0.061	0.058	0.054	0.052	0.049	0.046	0.044	0.042	0.040	0.038	0.036	0.035	0.033	0.031	0.031	0.029	0.028	0.027	0.025
		out	0.776	0.143	0.079	0.063	0.054	0.049	0.043	0.038	0.035	0.033	0.030	0.028	0.028	0.028	0.028	0.028	0.029	0.028	0.028	0.028	0.028	0.029
5-types	EC, IA, RD, LC, SN	in	0.000	0.067	0.066	0.061	0.058	0.054	0.052	0.049	0.046	0.044	0.042	0.040	0.038	0.036	0.035	0.033	0.031	0.031	0.029	0.028	0.027	0.025
		out	0.776	0.143	0.079	0.062	0.054	0.049	0.043	0.038	0.035	0.033	0.030	0.028	0.028	0.028	0.028	0.028	0.029	0.028	0.028	0.028	0.028	0.029

Table 2.D.5: In- and out-of-sample accuracy standard errors - Game B (continued)

			Decisions used to estimate and classify																					
			(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)	(31)	(32)	(33)	(34)	(35)	(36)	(37)	(38)	(39)	(40)	(41)	(42)	(43)	(44)
1-type	EC	in	0.057	0.055	0.052	0.050	0.048	0.045	0.043	0.040	0.038	0.036	0.034	0.032	0.030	0.029	0.027	0.025	0.022	0.020	0.018	0.016	0.013	0.042
		out	0.059	0.062	0.065	0.069	0.072	0.075	0.077	0.081	0.085	0.088	0.094	0.100	0.106	0.115	0.123	0.135	0.146	0.163	0.185	0.221	0.272	1.847
	IA	in	0.051	0.049	0.047	0.045	0.043	0.041	0.039	0.037	0.035	0.032	0.031	0.029	0.028	0.026	0.024	0.022	0.020	0.018	0.016	0.014	0.011	0.040
		out	0.052	0.054	0.056	0.060	0.062	0.065	0.067	0.070	0.075	0.078	0.083	0.087	0.094	0.101	0.107	0.116	0.127	0.141	0.158	0.188	0.231	1.524
	RD	in	0.051	0.049	0.047	0.045	0.043	0.041	0.039	0.037	0.036	0.033	0.032	0.030	0.029	0.027	0.025	0.023	0.021	0.020	0.017	0.015	0.012	0.040
		out	0.053	0.055	0.058	0.061	0.064	0.067	0.070	0.074	0.078	0.082	0.088	0.093	0.100	0.109	0.116	0.127	0.138	0.156	0.175	0.207	0.256	1.745
	LC	in	0.028	0.026	0.025	0.024	0.023	0.022	0.021	0.020	0.019	0.019	0.018	0.017	0.016	0.015	0.014	0.013	0.011	0.010	0.009	0.008	0.007	0.026
		out	0.033	0.033	0.033	0.035	0.036	0.037	0.039	0.041	0.043	0.046	0.048	0.050	0.054	0.058	0.062	0.066	0.073	0.081	0.088	0.108	0.138	0.921
	SN	in	0.024	0.023	0.022	0.021	0.020	0.019	0.018	0.018	0.017	0.016	0.015	0.014	0.014	0.013	0.012	0.011	0.010	0.009	0.008	0.007	0.007	0.022
		out	0.028	0.029	0.030	0.032	0.033	0.034	0.035	0.037	0.038	0.041	0.043	0.045	0.048	0.052	0.056	0.059	0.064	0.071	0.079	0.092	0.118	0.787
2-types	EC, IA	in	0.051	0.049	0.047	0.045	0.043	0.041	0.039	0.037	0.035	0.032	0.031	0.029	0.028	0.026	0.024	0.022	0.020	0.018	0.016	0.014	0.011	0.040
		out	0.051	0.054	0.056	0.060	0.062	0.064	0.067	0.070	0.074	0.078	0.083	0.087	0.094	0.101	0.107	0.116	0.127	0.141	0.158	0.188	0.231	1.524
	EC, RD	in	0.051	0.049	0.047	0.045	0.043	0.041	0.039	0.037	0.036	0.033	0.032	0.030	0.029	0.027	0.025	0.023	0.021	0.020	0.017	0.015	0.012	0.040
		out	0.053	0.055	0.058	0.061	0.064	0.067	0.070	0.074	0.078	0.082	0.088	0.093	0.100	0.109	0.116	0.127	0.138	0.156	0.175	0.207	0.256	1.745
	EC, LC	in	0.028	0.027	0.025	0.024	0.023	0.022	0.021	0.020	0.019	0.019	0.017	0.016	0.016	0.015	0.013	0.012	0.011	0.010	0.009	0.008	0.007	0.025
		out	0.032	0.033	0.033	0.035	0.035	0.037	0.039	0.041	0.043	0.046	0.048	0.050	0.054	0.058	0.062	0.066	0.072	0.080	0.088	0.106	0.135	0.917
	EC, SN	in	0.024	0.023	0.022	0.021	0.020	0.019	0.018	0.018	0.017	0.016	0.015	0.014	0.014	0.013	0.012	0.011	0.010	0.009	0.008	0.007	0.007	0.022
		out	0.028	0.029	0.030	0.032	0.033	0.034	0.035	0.037	0.038	0.041	0.043	0.045	0.048	0.052	0.056	0.059	0.064	0.071	0.079	0.092	0.118	0.787
	IA, RD	in	0.050	0.048	0.046	0.044	0.042	0.040	0.038	0.036	0.034	0.032	0.031	0.029	0.027	0.026	0.023	0.021	0.019	0.018	0.016	0.014	0.011	0.039
		out	0.050	0.053	0.055	0.058	0.060	0.063	0.065	0.069	0.073	0.076	0.081	0.086	0.092	0.099	0.105	0.113	0.125	0.139	0.156	0.186	0.229	1.521
	IA, LC	in	0.029	0.028	0.026	0.026	0.024	0.023	0.022	0.021	0.019	0.019	0.017	0.016	0.016	0.015	0.014	0.012	0.011	0.010	0.009	0.008	0.006	0.022
		out	0.034	0.034	0.036	0.037	0.038	0.040	0.041	0.043	0.045	0.048	0.050	0.053	0.057	0.061	0.065	0.069	0.076	0.083	0.091	0.109	0.138	0.914
	IA, SN	in	0.024	0.023	0.022	0.021	0.020	0.019	0.018	0.018	0.017	0.016	0.015	0.014	0.014	0.013	0.012	0.011	0.010	0.009	0.008	0.007	0.006	0.020
		out	0.028	0.029	0.030	0.031	0.032	0.034	0.035	0.037	0.039	0.041	0.043	0.045	0.049	0.052	0.055	0.058	0.063	0.070	0.078	0.093	0.118	0.778
	RD, LC	in	0.028	0.027	0.025	0.024	0.023	0.022	0.021	0.020	0.019	0.019	0.017	0.016	0.016	0.015	0.014	0.012	0.011	0.010	0.009	0.008	0.007	0.025
		out	0.032	0.033	0.033	0.035	0.036	0.038	0.040	0.041	0.043	0.047	0.048	0.051	0.055	0.058	0.062	0.066	0.073	0.081	0.089	0.108	0.136	0.929
	RD, SN	in	0.024	0.023	0.022	0.021	0.021	0.020	0.018	0.018	0.017	0.016	0.015	0.014	0.014	0.013	0.012	0.011	0.010	0.009	0.008	0.007	0.006	0.022
		out	0.029	0.030	0.031	0.032	0.033	0.034	0.035	0.037	0.039	0.041	0.043	0.045	0.048	0.052	0.056	0.059	0.064	0.072	0.079	0.093	0.118	0.784
	LC, SN	in	0.024	0.023	0.022	0.021	0.020	0.019	0.018	0.017	0.016	0.016	0.015	0.014	0.014	0.013	0.012	0.011	0.010	0.009	0.008	0.007	0.006	0.021
		out	0.028	0.030	0.030	0.032	0.032	0.033	0.034	0.036	0.037	0.039	0.041	0.043	0.046	0.050	0.053	0.057	0.062	0.069	0.076	0.090	0.114	0.743
3-types	EC, IA, RD	in	0.050	0.048	0.046	0.044	0.042	0.040	0.038	0.036	0.034	0.032	0.031	0.029	0.027	0.025	0.023	0.021	0.019	0.018	0.016	0.014	0.011	0.039
		out	0.050	0.053	0.055	0.058	0.060	0.063	0.065	0.069	0.073	0.076	0.081	0.086	0.092	0.099	0.105	0.113	0.125	0.139	0.156	0.186	0.229	1.521
	EC, IA, LC	in	0.029	0.028	0.026	0.026	0.024	0.023	0.022	0.021	0.019	0.019	0.017	0.016	0.016	0.015	0.014	0.012	0.011	0.010	0.009	0.008	0.006	0.022
		out	0.034	0.034	0.036	0.037	0.038	0.040	0.041	0.043	0.045	0.048	0.050	0.053	0.057	0.061	0.065	0.069	0.076	0.083	0.091	0.109	0.138	0.914
	EC, IA, SN	in	0.024	0.023	0.022	0.021	0.020	0.019	0.018	0.018	0.017	0.016	0.015	0.014	0.014	0.013	0.012	0.011	0.010	0.009	0.008	0.007	0.006	0.020
		out	0.028	0.029	0.030	0.031	0.032	0.034	0.035	0.037	0.039	0.041	0.043	0.045	0.049	0.052	0.055	0.058	0.063	0.070	0.078	0.093	0.118	0.778
	EC, RD, LC	in	0.027	0.025	0.025	0.024	0.023	0.022	0.021	0.020	0.019	0.018	0.017	0.016	0.016	0.015	0.013	0.012	0.011	0.010	0.009	0.008	0.007	0.025
		out	0.031	0.032	0.032	0.035	0.035	0.037	0.039	0.041	0.043	0.046	0.048	0.050	0.054	0.058	0.062	0.066	0.073	0.082	0.089	0.108	0.135	0.917
	EC, RD, SN	in	0.024	0.023	0.022	0.021	0.021	0.020	0.018	0.018	0.017	0.016	0.015	0.014	0.014	0.013	0.012	0.011	0.010	0.009	0.008	0.007	0.006	0.022
		out	0.029	0.030	0.031	0.032	0.033	0.034	0.035	0.037	0.039	0.041	0.043	0.045	0.048	0.052	0.056	0.059	0.064	0.072	0.079	0.093	0.118	0.784
	EC, LC, SN	in	0.024	0.023	0.022	0.021	0.020	0.019	0.018	0.017	0.016	0.016	0.015	0.014	0.014	0.013	0.012	0.011	0.010	0.009	0.008	0.007	0.006	0.021
		out	0.028	0.030	0.030	0.032	0.032	0.033	0.034	0.036	0.037	0.039	0.041	0.043	0.046	0.050	0.053	0.057	0.062	0.069	0.076	0.090	0.114	0.743
	IA, RD, LC	in	0.029	0.028	0.026	0.026	0.024	0.023	0.022	0.021	0.019	0.019	0.017	0.016	0.016	0.015	0.014	0.012	0.011	0.010	0.009	0.008	0.006	0.022
		out	0.034	0.034	0.036	0.037	0.038	0.040	0.041	0.043	0.045	0.048	0.050	0.053	0.057	0.061	0.065	0.069	0.076	0.083	0.091	0.109	0.138	0.914
	IA, RD, SN	in	0.024	0.023	0.022	0.021	0.020	0.019	0.018	0.018	0.017	0.016	0.015	0.014	0.014	0.013	0.012	0.011	0.010	0.009	0.008	0.007	0.006	0.020
		out	0.028	0.029	0.030	0.031	0.032	0.033	0.034	0.037	0.039	0.041	0.043	0.045	0.049	0.052	0.055	0.058	0.064	0.070	0.078	0.093	0.117	0.778
	IA, LC, SN	in	0.024	0.023	0.022	0.021	0.020	0.020	0.018	0.018	0.017	0.017	0.016	0.015	0.014	0.013	0.012	0.011	0.010	0.009	0.008	0.007	0.006	0.019
		out	0.030	0.031	0.032	0.033	0.033	0.035	0.036	0.038	0.040	0.041	0.044	0.046	0.050	0.053	0.056	0.060	0.065	0.072	0.079	0.095	0.119	0.783
	RD, LC, SN	in	0.024	0.023	0.022	0.021	0.020	0.019	0.018	0.017	0.016	0.016	0.015	0.014	0.014	0.013	0.012	0.011	0.010	0.009	0.008	0.007	0.006	0.021
		out	0.028	0.030	0.030	0.032	0.032	0.033	0.034	0.036	0.037	0.039	0.041	0.043	0.046	0.050	0.053	0.057	0.062	0.069	0.076	0.090	0.114	0.743
4-types	EC, IA, RD, LC	in	0.029	0.028	0.026	0.026	0.024	0.023	0.022	0.021	0.019	0.019	0.017	0.016	0.016	0.015	0.014	0.012	0.011	0.010	0.009	0.008	0.006	0.022
		out	0.034	0.034	0.036	0.037	0.038	0.040	0.041	0.043	0.045	0.048	0.050	0.053	0.057	0.061	0.065	0.069	0.076	0.083	0.091	0.109	0.138	0.914
	EC, IA, RD, SN	in	0.024	0.023	0.022	0.021	0.020	0.019	0.018	0.018	0.017	0.016	0.015	0.014	0.014	0.013	0.012	0.011	0.010	0.009	0.008	0.007	0.006	0.020
		out	0.028	0.029	0.030	0.031	0.032	0.033	0.034	0.037	0.039	0.041	0.043	0.045	0.049	0.052	0.055	0.058	0.064	0.070	0.078	0.093	0.117	0.778
	IA, RD, LC, SN	in	0.024	0.023	0.022	0.021	0.020	0.020	0.018	0.018	0.017	0.017	0.016	0.015	0.014	0.013	0.012	0.011	0.010	0.009	0.008	0.007	0.006	0.019
		out	0.030	0.031	0.032	0.033	0.033	0.035	0.036	0.038	0.040	0.041	0.044	0.046	0.050	0.053	0.056	0.060	0.065	0.072	0.079	0.095	0.119	0.783
	EC, RD, LC, SN	in	0.024	0.023	0.022	0.021	0.020	0.019	0.018	0.017	0.016	0.016	0.015	0.014	0.014	0.013	0.012	0.011	0.010	0.009	0.008	0.007	0.006	0.021
		out	0.028	0.030	0.030	0.032	0.032	0.033	0.034	0.036	0.037	0.039	0.041	0.043	0.046	0.050	0.053	0.057	0.062	0.069	0.076	0.090	0.114	0.743
	EC, IA, LC, SN	in	0.024	0.023	0.022	0.021	0.020	0.020	0.018	0.018	0.017	0.017	0.016	0.015	0.014	0.013	0.012	0.011	0.010	0.009	0.008	0.007	0.006	0.019
		out	0.030	0.031	0.032	0.033	0.033	0.035	0.036	0.038	0.040	0.041	0.044	0.046	0.050	0.053	0.056	0.060	0.065	0.072	0.079	0.095	0.119	0.783
5-types	EC, IA, RD, LC, SN	in	0.024	0.023	0.022	0.021	0.020	0.020	0.018	0.018	0.017	0.017	0.016	0.015	0.014	0.013	0.012	0.011	0.010	0.009	0.008	0.007	0.006	0.019
		out	0.030	0.031	0.032	0.033	0.033	0.035	0.036	0.038	0.040	0.041	0.044	0.046	0.050	0.053	0.056	0.060	0.065	0.072	0.079	0.095	0.119	0.783

This table presents the standard errors of the in-sample and out-of-sample accuracies of all 31 worlds in Game A. Standard errors are derived as $se = \sigma_{\mu_{\text{acc}}}/\sqrt{S}$, where $\sigma_{\mu_{\text{acc}}}$ refers to the standard deviation of the mean accuracy μ_{acc} which equals the average of the $M = 1000$ computed accuracies. The accuracy itself is derived as the ratio of correctly fitted decisions over the total number of decisions. We assume Senders' first-order beliefs to equal one. EC reflects the standard economic model, IA inequity aversion, RD reference-dependent preferences, LC lying costs, and SN reputation or social norms. Each column corresponds to a specific number of decisions used to estimate individual parameters, classify participants into types and to compute the in-sample performance. The remaining decisions are used to derive the out-of-sample fit. All standard errors are multiplied by 100.

3 Indebtedness, Interests, and Incentives: State-contingent Sovereign Debt Revisited*

with Felix Fattinger

3.1 Introduction

Even though the most acute symptoms of the Eurozone's sovereign debt crisis have abated, the necessary precondition for its recrudescence still persists to date: massive debt-to-GDP levels in many European countries. Surging public debt burdens arguably pose substantial refinancing risks, if perceived default probabilities should shift upwards again (Calvo (1988)). This has led both the press as well as renowned macroeconomists to recently rediscover the benefits of state-contingent bonds as an alternative refinancing instrument,[1] but this time for the pre-restructured debt of advanced economies.[2]

This paper investigates the welfare implications of state-contingent debt as a temporary device against swelling refinancing costs of highly indebted sovereigns. Our focus lies on governments whose economically required deleveraging efforts under standard sovereign debt are so immense that they have to be considered politically infeasible. We demonstrate state-contingent debt's virtue of providing more space for mitigating sovereign default risk. If sovereign borrowing is contractible—a requirement whose practicability is

* We would like to thank Markus Brunnermeier, Marc Chesney, Sebastian Dörr, Lucas Fuhrer, Seraina Grünewald, Michel Habib, David Hémous, Andreas Müller, Steven Ongena, Jean-Charles Rochet, Alexander Wagner, Jiri Woschitz, Alexandre Ziegler, Fabrizio Zilibotti, and participants at the UZH Finance Poster Workshop 2015, the Zurich Workshop on Economics 2015, the MFA 2016, the Spring Meeting of Young Economists 2016, and the URPP Financial Market Regulation doctoral colloquium at the University of Zurich for very helpful comments and suggestions. A version of this paper is being circulated as a working paper.

[1] For example, 'The Economist' referred to suggestions from Morgan Stanley to allow governments to issue GDP-linked bonds (source: A chronic problem, *The Economist*, May 14, 2016, `http://www.economist.com/node/21698669`). Blanchard et al. (2016) argue for a large scale issuance of growth-indexed bonds in advanced economies during normal times. They show for several Eurozone countries how GDP-indexed bonds could significantly reduce the upper-tail risk of excessive sovereign indebtedness, thereby attenuating implied default probabilities.

[2] Historically, securities linked to national growth have only been issued during the process of debt restructuring. More recently, Argentina (2005), Greece (2012), and Ukraine (2015) issued warrant-like instruments, where payments are conditional on realized growth rates.

discussed—this even holds true in the presence of low growth expectations, i.e., when the issuance of standard GDP-indexed bonds is particularly difficult.

In light of the European Central Bank (ECB)'s most effective crisis response, i.e., its 'Outright Monetary Transactions' program (OMT), we consider two types of state-contingent debt instruments: (i) puttable debt and (ii) GDP-to-debt-linked bonds, henceforth simply GDR (GDP-to-debt ratio) bonds. Investors in puttable debt receive the right to put their claims with a third party in return for a fixed payment and are thus insured against sovereign default. GDR bonds offer state-contingent interests that are *inversely* related to sovereign indebtedness. Hence, conditionally on sufficient deleveraging, their interest payments can still be positive under zero or even negative growth.

We find that, in the presence of non-negligible default costs, state-contingent bonds generally lead to substantial welfare gains. In terms of relative welfare improvements for risk-averse consumers, the *counter-cyclical* fiscal spending allowed for by GDR bonds' *cyclical* interest charges (as a function of GDP) proves superior to puttable bonds' simple insurance mechanism. For both instruments, governments deleveraging incentives as well as practical implementation issues are discussed.

The ECB's crisis mitigation efforts covered a wide range of non-standard monetary policy measures: from expanding volume and maturities of lending operations to adjusting collateral requirements for repo transactions, followed by direct interventions in securities markets. Whereas the effects of the first such direct intervention, i.e., the 'Securities Markets Program' (SMP), remained limited, the sole announcement of its successor OMT in July 2012 significantly calmed sovereign bond markets. In essence, OMT allows the ECB to buy unlimited amounts of distressed government bonds (with maturities up to three years) on the secondary market, as long as the issuing government agrees to comply with certain coercive measures.[3]

Following Merton (1977), OMT's conditional guarantee to buy traded sovereign bonds to reduce distressed yields can be interpreted as a written put option: The issuing government has the right to initiate purchases of its own bonds, once their price has fallen below a certain threshold. Since its first announcement, this ex-ante free option indirectly offered to OMT eligible countries received considerable opposition among Eurozone countries, particularly from Germany.[4] Concerns were raised that OMT's free option results in substantial risk redistribution among member countries and fails to sufficiently reinforce

[3] For an overview of the ECB's non-standard measures up to and including OMT see, e.g., the speech by Peter Praet, member of the executive board of the ECB, on April 17, 2013 (source: https://www.ecb.europa.eu/press/key/date/2013/html/sp130417.en.html).

[4] The Federal Constitutional Court of Germany questioned OMT's compliance with EU law. It was finally declared to be in line with the ECB's official mandate of price stability and thus considered legal by the European Court of Justice on June 16, 2015 (source: judgment of the court in case C-62/14, June 16, 2015, http://curia.europa.eu/jcms/jcms/j_6/).

fiscal discipline of candidate governments.

In the context of OMT, state-contingency is achieved through lower interest rates during times of high perceived default probabilities. The closest *incentive-compatible* alternative to OMT are puttable bonds, where an intergovernmental organization insures investors against default, thereby substantially lowering interest payable by the issuing government. Importantly, contrary to OMT, the issuing government has to *ex-ante* compensate the insurance provider by paying the insurance premium in return.[5] Thus, puttable bonds internalize the risk-sharing costs which remain non-remunerated under OMT.

First, motivated by OMT's inherent insurance mechanism, we start our analysis with puttable debt in the absence of default costs other than investors' foregone repayments, and relate our results to the seminal corporate debt valuation model by Merton (1974). The rationale underlying puttable sovereign debt is simple. Newly issued puttable bonds contain an embedded put option which serves as default protection for investors. If the sovereign is unable to refinance its liabilities at maturity, i.e., whenever the government defaults on its puttable debt, the writer of the embedded put option repays investors in full (principal and accrued interests) and acquires the initial claim against the issuing sovereign.

Abstracting from any credit risk on the insurance guarantor's side, puttable bonds correspond to a risk-free investment and therefore only pay the risk-free rate. In the context of OMT, default protection is provided by the ECB, i.e., more precisely by its stakeholders. In sharp contrast to OMT, in order to issue puttable debt, the sovereign would have to ex-ante compensate the ECB or another sufficiently capitalized intergovernmental agency for writing the embedded option.

Merton (1974) shows that the spread between risky debt and otherwise similar but credit risk-free debt can be interpreted as the value of a put option on total assets. In the absence of default costs, we demonstrate the validity of Merton's (1974) equivalence result within the sustainable sovereign debt model of Collard et al. (2015). They model sovereign borrowing to be solely constrained by credit markets' willingness to lend. Relying on uniformly distributed GDP growth rates, which allow for closed-form solutions, we show that a risk-neutral government is exactly indifferent between paying for protection against increased refinancing pressure today versus accepting lower publicly financed consumption or even the risk of default tomorrow. However, as soon as default imposes additional costs on the defaulting government, the lower default probability under puttable debt's reduced financing costs becomes welfare improving.

Second, we introduce costs incurred in case of sovereign default. As soon as default

[5] In contrast to credit default swaps (CDS) on sovereign debt, the embedded default insurance of puttable bonds can not be disentangled from the underlying default exposure. Moreover, the insurance premium is paid in full at issuance, rather then periodically over the bonds' lifetime.

imposes significant disutility on the defaulting government,[6] it has incentives not to overborrow. Based on actual country-specific yield spreads, we calibrate these costs such that the implied default probabilities are consistent with Merton's (1974) equivalence result. Hence, in the presence of non-negligible default costs, instead of following Collard et al. (2015), we adopt a one-shot perspective, where a risk-neutral government trades off publicly financed private consumption against expected default costs when choosing its optimal borrowing rate. In particular, our starting point is an already highly indebted government whose outstanding debt contains considerable default risk and therefore imposes high refinancing costs.[7]

Within this new setting, we compare utilities between alternative state-contingent refinancing instruments. In the spirit of Blanchard et al.'s (2016) current analysis for various Eurozone countries, we additionally consider GDP-linked debt. To allow state-contingent debt instruments to be welfare improving, we introduce a risk-averse consumer, who optimizes her utility from publicly financed consumption conditional on the government's borrowing policy. In order to account for the different frequencies between private consumption decisions and the passage of public budgets, the former is modeled in continuous time, whereas public borrowing remains constant over the considered period.

In contrast to standard GDP-linked debt, GDR bonds' interest payments *inversely* depend on a government's *relative indebtedness*, i.e., whenever its GDP-to-debt ratio rises, the payable interests increase. In addition to imposing state-contingent financing costs, GDR bonds provide a new signaling device for the issuing government. Even if credit markets have very pessimistic growth projections, GDR bonds can still offer competitive risk-return profiles to reluctant investors, *given* that the issuing government can credibly commit itself to a sustainable deleveraging. In case of insurmountable limited commitment concerns by potential investors, equipping GDR bonds with an embedded put option written by a third counterparty, i.e., similar to puttable bonds, could foster the enforceability of previous deleveraging commitments. The assertiveness of the former

[6] Examples comprise trade disruption or reputational costs for the defaulting government (see Section 3.3).

[7] In contrast to the analysis of puttable debt within the model of Collard et al. (2015), increased refinancing pressure here occurs due to high perceived default probabilities, instead of tighter credit market conditions. Lane (2012) provides an intuitive explanation, how—in the presence of high indebtedness—sudden shifts in sentiments can cause a multicountry currency union to jump from a sustainable equilibrium to one with highly increased yields and unsustainable debt levels. In the context of the Eurozone, such effects were likely amplified by internal flights-to-safety. Indeed, the empirical evidence of De Grauwe and Ji (2012) suggests a significant effect of negative market sentiments on the spreads of Greece, Ireland, Portugal, and Spain between 2010 and 2011. Up to the introduction of OMT, the severe effects of market sentiments were also pointed out by both the International Monetary Fund (IMF) (source: IMF Global Financial Stability Report, April 2012, http://www.imf.org/external/pubs/ft/gfsr/2012/01/) and the ECB (source: Draghi, M., ECB press conference, September 6, 2012, http://www.ecb.europa.eu/press/pressconf/2012/html/is120906.en.html).

European Troika in the Eurozone's recent debt renegotiations provides some evidence for such a supranational party's potential enforcement power (see discussion in Section 3.5).

We conduct a sensitivity analysis of optimal government borrowing under (i) standard sovereign debt, (ii) puttable debt, and (iii) GDR bonds. In the context of all three instruments, we confirm the model's capability to deliver intuitive results with respect to its parameters. For instance, we find the government's optimal borrowing rate to be increasing in GDP growth, U-shaped in growth volatility (manageable versus unmanageably high growth risk), and to decrease in consumer risk aversion and default costs. Moreover, GDR bonds' state-contingency reduces the interest burden during times of low growth. This allows the issuing government to deleverage less dramatically than under standard or puttable debt, while keeping its default probability constant.

Third, in order to evaluate the welfare implications of state-contingent debt in the presence of default costs, we run a case study calibrating our model to the five Eurozone countries whose bond yields were most heavily affected by the debt crisis: Portugal, Ireland, Italy, Greece, and Spain. We calibrate country-specific default costs to match historical default probabilities implied by sovereign credit ratings prior to the ECB's OMT announcement. We find that, across growth scenarios, switching from standard to puttable debt leads to welfare improvements.

In comparison to puttable debt, GDR bonds consistently yield superior utility levels. Contrary to former's simple insurance mechanism, GDR bonds' state-contingent interest charges allow the risk-averse representative agent to considerably smooth her within-period consumption path. Hence, in our calibration, GDR bonds' consumption smoothing effect outweighs puttable debt's more substantial reduction in expected default costs. Moreover, we show that even for very pessimistic growth projections (i.e., the fifth percentile of historical growth rates), GDR bonds' expected Sharpe ratios are around unity (or even higher) for all five countries.

Our paper relates to the literature on state-contingent sovereign bonds. Emerging economies regularly face the risk of exogenous and undiversifiable income shocks that may cause capital flow reversals and potentially lead to severe contractions. In order to provide developing countries the possibility to hedge themselves against non-contractible shocks, Caballero (2003) proposes the International Monetary Fund (IMF) to supply a market for state-contingent bonds. Such bonds would reduce refinancing pressure in case of a severe income shock by, e.g., relating interest payments to income flows or by offering state-contingent insurance payments. As an example for the latter, Caballero (2003) argues that "Chile could eliminate most, if not all, of its deep recessions by embedding into its external bonds a long-term put option, yielding US$ 6-8 billions when the price of copper [its main export commodity] falls by more than two standard deviations" (Caballero, 2003, p. 34).[8] Caballero's (2003) proposal can be viewed as an alternative to the IMF's already existing contingent credit lines for emerging economies. The potentially stabilizing properties of puttable sovereign bonds for emerging countries without access to interest rate derivative markets are pointed out by Neftci and Santos (2003).

GDP-linked securities represent the most prominent example of state-contingent sovereign debt discussed in the literature. Shiller (1994, 2003) argues for macro markets to trade GDP-linked perpetual claims on fractions of countries' respective GDP. Borensztein et al. (2004) identify four major benefits of GDP-linked bonds over standard sovereign bonds: (i) lower likelihood of sovereign defaults, (ii) fewer damaging pro-cyclical fiscal policy implementations, (iii) the opportunity for a smoother intertemporal tax path, and (iv) lower likelihood of government overspending during booms.[9] In the context of the Eurozone's recent debt crisis, the stabilizing effects of GDP-linked bonds on the public finances of advanced economies have repeatedly been emphasized by Barr et al. (2014) and Blanchard et al. (2016), among others.

Naturally, our paper also relates to the numerous alternative proposals brought forward in response to the currency union's vulnerability to surging public debt levels. The so-called 'Eurobonds', probably the most prominently discussed of all suggestions, refer to government bonds jointly issued by all Eurozone member countries. Since their credit risk relies on the solvency of the Eurozone as a whole, the associated reduction in the risk of destabilizing attacks on national government bond markets would be substantial (Favero and Missale (2012)).

The general concern among fiscally stronger member states (Germany in particular) has been that Eurobonds could incentivize less disciplined member states to overborrow. Hellwig and Philippon (2011) share this concern and consequently propose the introduction of common senior debt with maturity of less than one year called 'Eurobills' aiming at easening debt rollover for Eurozone countries. Eurobills would be introduced permanently and participation in the market tied to compliance with certain governance criteria. Brunnermeier et al. (2011) suggest the creation of a 'European Debt Agency' that would buy national sovereign bonds up to 60% of GDP for each member country. By securitization and the re-issuance of different tranches of bonds, it could satisfy the demand for a safe asset which should also be used as main collateral in central bank liquidity operations.

The remainder of the paper is organized as follows. Section 3.2 analyzes the effects of puttable debt for a shortsighted risk-neutral government in the absence of default costs.

---

[8] In the context of private corporations, Chidambaran et al. (2001) discuss how the gold-mining company Freeport McMoRan's usage of gold-linked depository shares helped to reduce its financing costs.

[9] For a detailed literature overview on GDP-linked bonds see, e.g., Barr et al. (2014).

While allowing for default costs, Section 3.3 additionally introduces GDR bonds in the context of a continuously consuming risk-averse agent. In Section 3.4, relative welfare effects of both puttable and GDR bonds relative to standard sovereign debt are estimated based on an empirical case study. Section 3.5 discusses potential implementation issues of GDR bonds and presents our concluding remarks.

3.2 Puttable Debt in the Absence of Default Costs

Motivated by OMT's implicitly written put option, we begin our analysis of state-contingent debt by introducing puttable bonds into a stylized model of sovereign borrowing, where default does not cause any costs to the defaulting government. We deliberately choose this starting point, as the absence of default costs allows for a useful benchmark assessment of puttable bonds' potential welfare effects.

In summary, if costs of default are negligible, Merton's (1974) reinterpretation of risk-adjusted spreads as put options can be explicitly verified within Collard et al.'s (2015) closed model of endogenous government borrowing. This insight proves valuable when evaluating puttable debt's welfare enhancing reduction in default risk in the presence of nonzero default costs as introduced in Section 3.3.

Abstracting from any kind of default costs is similar to the assumption of a short-sighted government discussed in Rochet (2006) and Collard et al. (2015).[10] Shortsighted governments always borrow as much as possible, since they do not take future repayments into account. Under this extreme assumption, sovereign debt levels are capped *only* by credit markets' willingness to lend.

Furthermore, we start by assuming lenders to be shortsighted as well, i.e., in the sense that interest spikes from the distant past do not make them anticipate any potential tightening of credit availability in the future. In other words, until confronted by a credit shock, they regard the risk-free interest rate as constant. This assumption is not crucial, but considerably simplifies the model's introduction, and is subsequently relaxed when we analyze puttable debt as response to increasing refinancing costs.

3.2.1 Government Borrowing

Relying on Collard et al.'s (2015) notation, we denote by Y_t the country's GDP at period t, by b_t the incumbent government's date-t proceeds from issuing zero-coupon debt with face value d_t maturing at $t+1$, both expressed as fractions (denoted in lowercase) of Y_t,

[10] Collard et al. (2016) argue that the therefrom deduced concept of 'excusable default'—governments only default on maturing debt if their budget constraint does not allow them to repay—better explains empirically observed sovereign debt levels than the more familiar concept of 'strategic default'.

and by α the government's primary surplus, also expressed as a fraction of Y_t and assumed to be constant over time. At any given date t, publicly financed consumption c_t is given by

$$c_t = ((1-\alpha) + b_t)Y_t - d_{t-1}Y_{t-1}, \tag{3.1}$$

i.e., private consumers receive the sum of their income net of the government's surplus and new borrowing proceeds net of expenses from repaying maturing government debt. Intuitively, sovereign default then occurs at $t+1$ if

$$(\alpha + b_{t+1})Y_{t+1} < d_t Y_t, \tag{3.2}$$

i.e., if the sum of primary surplus and proceeds from newly issued debt are insufficient to repay its current creditors.[11] We assume that even in case of default, publicly financed consumption is never negative, i.e., $c_t \geq 0$.

Collard et al. (2015) assume zero recovery in default.[12] Given its shortsightedness, the government aims to maximize the proceeds from borrowing at every date t. Hence, its t-proceeds from borrowing are then given by

$$b_t = \max_{d_t} \frac{d_t}{1+R_t} \mathbb{P}((\alpha + b_{t+1})Y_{t+1} \geq d_t Y_t), \tag{3.3}$$

where R_t denotes the market-implied risk-adjusted discount rate prevailing at t. Eq. (3.3) simply states that investors are only willing to lend the discounted expected repayment value of the newly issued zero-coupon bond. Assuming risk-neutral and shortsighted credit markets, Eq. (3.3) can be written as

$$b_t = \max_{d_t} \frac{d_t}{1+r} \mathbb{P}((\alpha + b_{t+1})Y_{t+1} \geq d_t Y_t), \tag{3.4}$$

where r refers to the risk-free interest rate, initially regarded as time-invariant by lenders.

The right-hand side of Eq. (3.4) exhibits a 'Laffer curve property'. At the maximum, the effects of increasing debt levels and default probabilities offset each other. Cond. (3.2) is equivalent to

$$(\alpha + b_{t+1})g_{t+1} < d_t,$$

where $g_{t+1} \equiv Y_{t+1}/Y_t$ denotes the continuous growth rate between date t and $t+1$ assumed

[11] This condition of sovereign default is different from Eaton and Gersovitz' (1981) concept of strategic default. Collard et al. (2015) refer to mounting empricial evidence that most sovereigns do not default voluntarily. Moreover, Collard et al. (2016) show that calibrated 'excusable' default models, i.e., in the spirit of Cond. (3.2), yield much more realistic debt-levels than models of strategic default.

[12] Alternatively, one could assume that creditors can (partially) claim the government's primary surplus upon default. However, even if creditors receive the entire primary surplus, the maximum sustainable debt in Collard et al. (2015) only increases slightly, due to its very low associated default probability.

$$
\begin{tikzcd}
 & & & r_t = r_H \arrow[r, dashed, "\pi_H"] \arrow[dr, dashed, "\pi_L"] & r_{t+1} = r_H \\
\cdots\cdots \arrow[r] & r_{t-2} = r_L \arrow[r] & r_{t-1} = r_L \arrow[ur] & & r_{t+1} = r_L
\end{tikzcd}
$$

Figure 3.1: Unexpected rise in risk-free interest rates

Notes: This figure illustrates a rare and unanticipated tightening of credit market conditions at t, modeled by an increase in the risk-free interest rate r_t from r_L to r_H. Thereafter, credit markets become aware of interest rates' stochastic nature and form beliefs about future risk-free rates at $t+1$: either r_{t+1} stays at r_H with probability π_H, or it decreases to r_L with probability $\pi_L = 1 - \pi_H$.

to be i.i.d. with cdf $F(\cdot)$. Thus, relying on future borrowing proceeds b_{t+1}, the government avoids default at $t+1$ with probability

$$\mathbb{P}((\alpha + b_{t+1})g_{t+1} \geq d_t) = \mathbb{P}\left(g_{t+1} \geq \frac{d_t}{\alpha + b_{t+1}}\right) \equiv 1 - F\left(\frac{d_t}{\alpha + b_{t+1}}\right).$$

Collard et al. (2015) define maximum sustainable borrowing b_M as the fixed point

$$b_M = \frac{\alpha + b_M}{1 + r} g_M(1 - F(g_M)), \tag{3.5}$$

where

$$(\alpha + b_M)g_M \equiv d_M$$

identifies maximum sustainable debt d_M and g_M satisfies

$$g_M = \arg\max_g g(1 - F(g)). \tag{3.6}$$

We now introduce a credit shock scenario as presented in Figure 3.1. After many consecutive periods with low risk-free rates r_L, an unanticipated shock at t tightens credit availability. This has two immediate effects: First, the risk-free rate rises from r_L to r_H. Second, shortsighted lenders are reminded of interests' time variability, which directly affects their willingness to lend, i.e., Eq. (3.4) becomes

$$b_t = \max_{d_t} \frac{d_t}{1 + r_H} \mathbb{P}((\alpha + b_{t+1})Y_{t+1} \geq d_t Y_t), \tag{3.7}$$

i.e., the discount rate increases to r_H.

Given the present interest rate shock, lenders *now* consider interest rates to be random but stationary, i.e.,

$$r_s \overset{\text{i.i.d.}}{\sim} r \sim r_L + B(1, \pi_H)(r_H - r_L) \;\; \forall s \geq t+1, \tag{3.8}$$

where $B(1, \pi_H)$ denotes a Bernoulli distribution with 'success' probability π_H. Defining $\beta_t \equiv (1 + r_t)b_t$ and recalling the stationarity of the country's growth rate, i.e., $g_s \overset{\text{i.i.d.}}{\sim} g \; \forall s \geq t + 1$, we can rewrite Eq. (3.7) as

$$\beta_t = \max_{d_t} d_t \mathbb{P}\left(\left(\alpha + \frac{\beta_{t+1}}{1+r}\right) g \geq d_t\right).$$

With stochastic interest rates as defined in Eq. (3.8), maximum sustainable debt d_M becomes a function of β_M given by

$$d_M(\beta_M) = \arg\max_d d\left(1 - \pi_L \underbrace{F\left(\frac{d}{\alpha + \frac{\beta_M}{1+r_L}}\right)}_{\substack{\text{default probability} \\ \text{given } d \text{ if } r = r_L}} - \pi_H \underbrace{F\left(\frac{d}{\alpha + \frac{\beta_M}{1+r_H}}\right)}_{\substack{\text{default probability} \\ \text{given } d \text{ if } r = r_H}}\right), \tag{3.9}$$

where maximum sustainable borrowing $b_M(t) = \beta_M/(1 + r_t)$ is state-dependent and β_M is the corresponding fixed point, i.e.,

$$\beta_M = d_M(\beta_M)\left(1 - \pi_L F\left(\frac{d_M(\beta_M)}{\alpha + \frac{\beta_M}{1+r_L}}\right) - \pi_H F\left(\frac{d_M(\beta_M)}{\alpha + \frac{\beta_M}{1+r_H}}\right)\right). \tag{3.10}$$

Due to the uncertainty about future interest rates, maximum sustainable debt in Eq. (3.9) is no longer indirectly implied by g_M in Eq. (3.6), but instead needs to take into account r's binomial distribution and the state-dependent default thresholds.

As demonstrated by Rochet (2006), assuming g to be uniformly distributed on $[1 + \mu_g - \sigma_g, 1 + \mu_g + \sigma_g]$, Eq. (3.9) and Eq. (3.10) then imply

$$d_M = \frac{1 + \mu_g + \sigma_g}{2\left(\frac{\pi_L}{\alpha + \frac{\beta_M}{1+r_L}} + \frac{\pi_H}{\alpha + \frac{\beta_M}{1+r_H}}\right)} \tag{3.11}$$

$$\beta_M = \frac{(1 + \mu_g + \sigma_g)^2}{8\sigma\left(\frac{\pi_L}{\alpha + \frac{\beta_M}{1+r_L}} + \frac{\pi_H}{\alpha + \frac{\beta_M}{1+r_H}}\right)}. \tag{3.12}$$

The following lemma gives the closed form solution to Eq. (3.12), on which we rely for our analysis of puttable debt below.

Lemma 1. *If $(r_L, r_H) \in [0, 1) \times [0, 1)$, $g \sim U[1 + \mu_g - \sigma_g, 1 + \mu_g + \sigma_g]$, and $(1 + \mu_g + \sigma_g)^2(8\sigma)^{-1} < \mathbb{E}[1 + r]$, then the fixed point in Eq. (3.12) is given by*

$$\beta_M = \frac{-B - \sqrt{B^2 - 4AC}}{2A} > 0,$$

where

$$A := \frac{(1+\mu_g+\sigma_g)^2}{8\sigma} - \mathbb{E}[1+r]$$
$$B := \alpha\left(\frac{(1+\mu_g+\sigma_g)^2}{8\sigma}(2+r_L+r_H) - (1+r_L)(1+r_H)\right)$$
$$C := \alpha^2\frac{(1+\mu_g+\sigma_g)^2}{8\sigma}(1+r_L)(1+r_H).$$

Proof. For proof see Appendix 3.A. □

3.2.2 Refinancing Costs During Credit Shock: Standard versus Puttable Debt

Suppose now that the government is given the possibility to buy protection against increased refinancing costs at t by issuing puttable instead of standard debt. The main difference between puttable and standard debt is that the former insures investors against sovereign default, i.e., in case of default at $t+1$, a third party guarantees to repay them in full. From investors' perspective, assuming the insurance writer to be sufficiently solvent, such puttable bonds correspond to a risk-free investment. We impose the following conditions for issuing puttable debt at t:

C1. The government's date-t proceeds may not surpass $b_M(t)$, i.e., they may not be higher than in the absence of puttable bonds.

C2. The government is required to ex-ante compensate the guarantor for the *initial* default risk taken over, i.e., for the default risk in absence of puttable bonds.

C3. The government is only allowed to issue puttable bonds, if the payment of the ex-ante insurance premium does not cause it to already default at t.

Condition C1 is intuitive given the intended risk managing character of puttable debt. Hence, it would be counterproductive, if sovereigns would be allowed to borrow more in the presence of high interest rates than available under low refinancing costs. In fact, C1 prevents higher borrowing due to a risk-shifting effect. C2 directly manifests itself in the ex-ante payable insurance premium (see below). Finally, C3 is simply imposed by the government's budget constraint which does not allow for negative consumption *after* borrowing proceeds.

Table 3.1 contrasts the government's two refinancing options at date t in more detail. For the issuing government, the refinancing costs of standard and puttable debt differ in two dimensions: on the one hand, the issuance of puttable bonds requires an ex-ante

Table 3.1: Standard vs. puttable debt

	Standard debt	Puttable debt
Type	zero-coupon	zero-coupon
Insurance	no	yes
Ex-ante premium	none	put price p_t
Spread	$R_t \equiv \frac{d_M}{b_M(t)} - 1$	risk-free rate r_t

Notes: This table compares standard and puttable debt available to the government to refinance its debt at t. R_t denotes the spread on date-t borrowing proceeds $b_M(t)$ to be paid to investors at $t+1$, if the government chooses standard debt. In case it chooses to issue puttable debt, p_t denotes the ex-ante premium to be paid by the issuing government to the insurance provider at t.

payment to the insurance guarantor at t. This premium compensates the third party for providing investors with insurance against sovereign default. The date-t value of the insurance p_t corresponds to the price of an European put option on sovereign debt with maturity $t+1$ and a strike price equal to the debt's face value plus accrued interests. On the other hand, the issuing government is in return protected against higher interests on its date-t borrowing proceeds due at $t+1$. Since investors are insured against default risk, the payable interests on $b_M(t)Y_t$ are reduced from the default-adjusted spread R_H to the current risk-free rate r_H, where

$$R_H \equiv \frac{d_M}{b_M|R_H} - 1,$$

and $b_M|R_H$ denotes maximum sustainable borrowing under high (default adjusted) interest rates.

Moving to an *intertemporal* perspective, the discounted expected sum of current and future publicly financed consumption equals

$$\mathbb{E}\left[\sum_{s\geq t}\delta^{s-t}c_s\right], \tag{3.13}$$

where δ is the discount factor and c_s is given by Eq. (3.1). Note that, for $\delta < 1$, the borrowing policy in Eq. (3.10) is consistent with the maximization of Eq. (3.13) by a risk-neutral government who is indifferent towards variations in private consumption. A risk-averse government, in contrast, might have an incentive to smooth private consumption across time by borrowing less if production is very high in order to avoid future default. In Section 3.3, we account for such consumption smoothing benefits by introducing a risk-averse consumer.

Table 3.2: Consumption under standard vs. puttable debt (no default costs)

	Standard debt (1)	Puttable debt (2)	Δ (1)−(2)
Consumption at t	c_t	$c_t - p_t$	p_t
Discounted expected consumption at $t+1$	$\delta\mathbb{E}[c_{t+1}\|R_H]$	$\delta\mathbb{E}[c_{t+1}\|r_H]$	$\delta b_M(t)\,(r_H - R_H)\,Y_t$

Notes: This table compares current and future publicly financed consumption levels under the issuance of standard and puttable debt, given an increase in the risk-free rate from r_L to r_H at t. The third column lists the differences at t and $t+1$ from the perspective of a risk-neutral government. p_t denotes the date-t put price, δ the government's discount factor, $b_M(t)$ date-t maximum borrowing under stochastic interest rates, and R_H the default-adjusted spread based on r_H.

Consistent with the above analysis, we can rewrite date-t private consumption as

$$c_t = ((1-\alpha) + b_M(t))Y_t - d_{t-1}Y_{t-1},$$

and at $t+1$, given a preceding refinancing with standard debt, as

$$c_{t+1}|R_H = ((1-\alpha) + b_M(t+1))Y_{t+1} - b_M(t)(1+R_H)Y_t,$$

where we rely on the identity $b_t(1+R_t) \equiv d_t$ with $R \in \{R_L, R_H\}$ denoting the state-dependent spread to be paid by the government on its borrowing proceeds.

In the presence of puttable debt, refinancing public debt with puttable bonds reduces date-t private consumption by the put option premium, i.e.,

$$c_t - p_t,$$

where at $t+1$ the put option either pays $b_M(t)(1+R_H)Y_t$ in case of default (zero recovery) or nothing otherwise. Note that the put option's strike price is set equal to the borrowing proceeds' face value plus accrued interests in absence of puttable debt (see C2 above). At $t+1$, due to the lower spread charged by insured investors, private consumption increases to

$$c_{t+1}|r_H = ((1-\alpha) + b_M(t+1))Y_{t+1} - b_M(t)(1+r_H)Y_t,$$

where it holds that $c_{t+1}|r_H \geq c_{t+1}|R_H$. The (discounted) differences in (expected) private consumption from issuing either standard or puttable debt are summarized in Table 3.2.

Lemma 2. *If $g \sim U[1+\mu_g - \sigma_g, 1+\mu_g+\sigma_g]$, then a risk-neutral government facing no default costs is indifferent between issuing standard or puttable debt in response to a credit* *shock.*

Proof. For proof see Appendix 3.A. □

Lemma 2 is in line with Merton (1974), implying that the put option price p_t is equal to the discounted spread $\delta(R_H - r_H)$ to be paid on the borrowing proceeds $b_M(t)Y_t$ financed by puttable bonds. In other words, a risk-neutral government's disutility from having to pay the insurance premium at t is exactly offset by higher private consumption due to lower interest payments at $t+1$. Hence, as long as there are no default costs, a risk-neutral government is indifferent between issuing standard or puttable debt in response to an increase in interest rates.

Proposition 1. *If $R_H > r_H$ and the probability of default under standard debt with spread R_H is nonzero, then the issuance of puttable debt always decreases the risk of sovereign default. In particular, if $g \sim U[1+\mu_g-\sigma_g, 1+\mu_g+\sigma_g]$ and $b_M(t)(1+R_H)/(\alpha+b_M(t+1)) > 1+\mu_g-\sigma_g$, then moving from standard to puttable debt decreases the government's default probability by*

$$\min\left(\frac{b_M(t)(R_H - r_H)}{2\sigma_g\left(\alpha + b_M(t+1)\right)}, \frac{\frac{b_M(t)(1+R_H)}{\alpha+b_M(t+1)} - (1+\mu_g-\sigma_g)}{2\sigma_g}\right),$$

where the first (second) term applies in case of a nonzero (zero) default probability under r_H.

Proof. For proof see Appendix 3.A. □

So far we have been abstracting from any costs to be borne by a defaulting sovereign. However, as soon as we introduce such sovereign default costs, Proposition 1 has important implications: reducing a government's default probability attenuates expected disutilities from default and, therefore, enhances social welfare. Hence, facing non-negligible default costs, even a risk-neutral government may be better off provided access to puttable bonds when facing increasing refinancing costs.

Furthermore, in the presence of a risk-averse consumer, a comparison to the consumption-smoothing benefits from GDP-linked bonds springs to mind. How puttable debt's reduction in expected default costs compares to GDP-linked debt's decrease in consumption variability is the focus of the following two sections.

3.3 State-contingent Borrowing in the Presence of Default Costs

Our previous working assumption of zero default costs incurred by a defaulting sovereign is arguably unrealistic. Borensztein and Panizza (2009) distinguish among four different types of sovereign default costs: (i) reputational costs, (ii) international trade exclusion costs, (iii) costs due to negative shocks on the domestic banking system, and (iv) political costs borne by the incumbent government. They find that the economic costs are generally significant, but short-lived. However, sovereign defaults often bring far-reaching consequences for elected officials. The authors document that in 18 out of 19 cases studied, the ruling coalition lost votes following the default and their electoral support declined on average by 16%.

Probably more meaningful to our case is the indirect measure of default costs given by the sheer amount of fiscal tightening that has been accepted by the Hellenic government during its ongoing debt negotiations in return for numerous bailout packages. In March 2012, the Greek government and its counter parties, i.e., the European Commission, Eurogroup, ECB, and IMF, signed the 'Second Economic Adjustment Programme for Greece'.[13] The austerity measures imposed by this second bailout package revised the total amount of fiscal cost reductions to approximately 65 billion Euros between 2010 and 2014, which corresponded to more than 30% of Greece's then GDP.[14]

As soon as default imposes significant disutility on the defaulting government, the latter has incentives not to overborrow. Based on actual country-specific yield spreads, we calibrate these costs such that the implied default probabilities are in line with Merton's (1974) equivalence result. Instead of following Collard et al. (2015), we adopt a simpler one-shot perspective, where a risk-neutral government trades off publicly financed private consumption against expected default costs when choosing its optimal borrowing rate. Given the intended temporary nature of the hereafter considered refinancing instruments, we deem such a one-period view appropriate.[15] In particular, our starting point is an already highly indebted government whose bonds contain a considerable default risk and therefore impose high refinancing costs. In contrast to the analysis of puttable debt within the model of Collard et al. (2015), increased refinancing pressure here occurs due to high

---

[13] The second package was followed by a third bailout package signed in 2015.

[14] Source: Excessive austerity killing Greece, *Kathimerini (English edition)*, September 30, 2012, `http://www.ekathimerini.com/145032/article/ekathimerini/business/excessive-austerity-killing-greece`.

[15] In addition, it allows us to calibrate the model to data from highly indebted Eurozone countries in Section 3.4. A model based on Collard et al. (2015) could not account for realistic default probabilities given the historically low GDP growth volatilities.

perceived default probabilities, instead of tighter credit market conditions.

Within this new setting, we compare utility gains from state-contingent refinancing instruments. In particular, similar to Blanchard et al. (2016), we now also consider GDP-linked debt. To allow such debt instruments to be welfare improving, we introduce a risk-averse consumer who optimizes her utility from publicly financed consumption conditional on the government's borrowing policy. In order to account for the time discrepancy between private consumption decisions and changes to public budget plans, the former are modeled continuously. Solving for private consumption in continuous time also allows us to consider the imperfect foreseeability of future debt levels due to deviations of private consumption from expectations, while fully accounting for GDP-linked debt's intertemporal consumption smoothing effect.

In contrast to standard GDP-linked bonds, we introduce bonds linked to a country's GDP-to-debt ratio, referred to as GDR bonds, where interest payments *inversely* depend on a government's *relative indebtedness*, i.e., implying counter-cyclical interest rate dynamics with respect to relative indebtedness. If a country's relative indebtedness increases, interest payments on GDR bonds decrease and vice versa. Thanks to GDR bonds' variable interest rates, negative shocks in GDP lead to a lower interest burden, relaxing the sovereign's refinancing pressure. In addition to imposing state-contingent financing costs, GDR bonds provide a new signaling device for the issuing government. Even if credit markets have very pessimistic growth projections, GDR bonds can still offer competitive risk-return profiles to reluctant investors, assuming that the issuing government can credibly commit itself to a sufficient deleveraging.[16]

Figure 3.2 summarizes the timing of our model where the government chooses its borrowing rate at t_0 which then remains fixed until $t_0 + 1$. During the period, a risk-averse consumer continuously maximizes her utility from publicly financed consumption, conditional on the government's fixed borrowing policy. At $t_0 + 1$, the government either refinances its debt (and sets a new borrowing rate) or defaults.

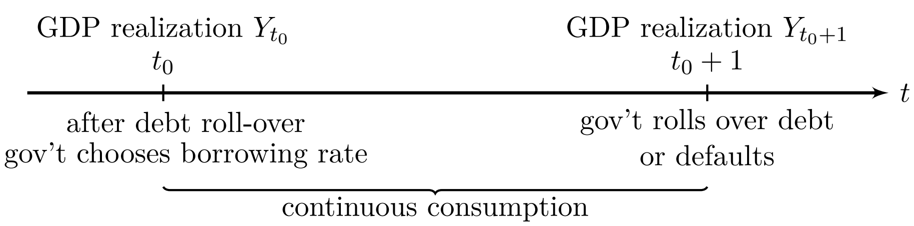


Figure 3.2: Timing of debt refinancing with continuous consumption

[16] In case of irreconcilable limited commitment concerns of potential investors, equipping the GDR bonds with a puttable component could foster the enforceability of initial deleveraging commitments. See also Section 3.5.

3.3.1 Dynamics - Indebtedness and GDP

The state variable of interest is a sovereign's annual GDP in relation to its accumulated public debt.[17] More specifically, we assume the following dynamics

$$dY_t = Y_t \left(\mu_Y dt + \sigma_Y dB_t \right), \tag{3.14}$$

where Y_t denotes the country's rolling annual GDP and B_t a standard Brownian motion. In addition, we model the dynamics of sovereign debt as

$$dD_t = D_t \mu_D dt, \tag{3.15}$$

where D_t corresponds to government date-t debt. Since we are considering a short-term decision problem (one-period borrowing under the risk of default), we consider it sensible to model the sovereign's real GDP growth rate μ_Y and GDP volatility σ_Y to be exogenous and constant over the considered time period. The proposed dynamics are consistent with Collard et al. (2015), who assume log-normally distributed growth rates. At the beginning of the period, i.e., at t_0, a country's government chooses its borrowing rate μ_D subject to its budget constraint.

We assume a country's short-term wealth W_t disposable for consumption to be additive in Y_t and D_t,[18] i.e.,

$$dW_t := (1 - \alpha)\, dY_t + (\mu_D - R)\, D_t dt - c_t dt, \quad W_{t_0} := (1 - \alpha) Y_{t_0} + D_{t_0}. \tag{3.16}$$

Here, c_t denotes the risk-averse consumer's continuous consumption and, as in Section 3.2, α indicates the country's primary surplus. Interest payments on government debt are given by RD_t, where R denotes the risk-adjusted interest rate, which is assumed constant over one period. Note that, in contrast to Section 3.2, R is now exogenous.

Under puttable government debt as described in Section 3.2, investors possess the option to put their bonds in case the country's debt-to-GDP level exceeds a certain threshold, i.e., if the country defaults. This credit insurance is provided by an intergovernmental agency which, in return, is compensated with the corresponding insurance premium (payable ex-ante). As a result, the issuing government only pays the risk-free interest on its debt but has to pay a fixed fee to the insurance guarantor at issuance. The budget

[17] This is in analogy to the third criterion of the euro convergence criteria (also known as the Maastricht criteria) which were established in 1992.

[18] As in Section 3.2, we deliberately abstract from consumption financed by issuing private debt.

constraint then reads

$$dW_t := (1-\alpha)\, dY_t + (\mu_D - r)\, D_t dt - c_t dt, \quad W_{t_0} := (1-\alpha)Y_{t_0} + D_{t_0} - p_{t_0}, \qquad (3.17)$$

where r denotes the constant risk-free interest rate and p_{t_0} corresponds to the put price at t_0, i.e., the ex-ante payable insurance premium, which lowers the initial wealth available for consumption.

The put price corresponds to the discounted value of the issued government debt multiplied by the probability of default. Assuming the insurance guarantor to be risk-neutral, the corresponding discount rate equals the risk-free rate.[19] The resulting put price is then given by

$$p_{t_0} = e^{-r} D_{t_0} \mathbb{P}\left(\frac{D_{t_0+1} - (W_{t_0+1} - W_{t_0})}{Y_{t_0+1}} - \frac{D_{t_0}}{Y_{t_0}} > 0 \right), \qquad (3.18)$$

where we impose, consistent with Section 3.2, zero recovery in default. For simplification, we assume that the country defaults whenever sovereign debt minus the change in disposable wealth relative to GDP is larger than the initial debt-to-GDP ratio. However, this somewhat strict default condition only has a small level effect and does not change our qualitative results. Moreover, in Section 3.4, we calibrate default costs such that the thereby induced default probability *under* the optimal borrowing rate is consistent with credit-rating-implied default risk. Since the debt-to-GDP ratio itself neglects changes in W_t between t_0 and t_0+1, we have to account for potential differences in disposable wealth due to the consumer's actual consumption. If she overconsumes, i.e., if she consumes more than production growth and public borrowing allows for, disposable wealth is reduced, which in turn increases the probability of sovereign default. We implicitly assume that all produced goods need to be consumed within one period and cannot be stored for later times.[20]

3.3.2 Government Borrowing with GDR Bonds and Default Costs

The government chooses its optimal borrowing rate subject to maximizing expected utility from consumption of its risk-averse consumer, while simultaneously accounting for

[19] Note that, in order to compute the put price, we consider that the government needs to pay the ex-ante risk-adjusted interest rate on its debt and neglect a lower interest rate's decreasing effect on the ex-post default probability. This is consistent with condition C2 in Section 3.2. Consequently, our estimated put price reflects a conservatively high estimate, as issuing puttable debt arguably reduces the chance that the put is exercised. Hence, a lower put price could possibly be negotiated, unless the intergovernmental agency has full bargaining power.

[20] In addition, we neglect investments which Gali et al. (2007) have shown to remain unaffected by a shock in government spending.

default risk. We solve the government's optimization problem by backward induction. First, following Merton's (1969) lifetime portfolio allocation approach, we determine the consumer's optimal consumption over time for a given constant borrowing rate μ_D set by the government. We model the risk-averse consumer as a representative agent with CRRA-utility. Given the borrowing rate μ_D, our representative agent chooses her optimal consumption plan $c_t^\star$ over a finite horizon, i.e., $\forall t \in (t_0, t_0+1)$, subject to her budget constraint. While doing so, she may consume more or less aggressively, depending on her time preference as well as the relative value she assigns to her end of period wealth. In addition, we assume that the representative agent, in contrast to the government, does not incorporate any default costs into her consumption decisions.

Let the dynamics of short-term wealth W_t disposable for consumption under GDR bonds be

$$dW_t := (1-\alpha)\, dY_t + \left(\mu_D dt - \frac{d\,(Y_t/D_t)}{Y_t/D_t}\right) D_t - c_t dt, \quad W_{t_0} := (1-\alpha)Y_{t_0} + D_{t_0}, \tag{3.19}$$

where $d\,(Y_t/D_t)\,/\,(Y_t/D_t)$ reflects the interest payments on GDR bonds, indicating the link between interest rates and relative indebtedness. The interest payments are increasing in changes in Y_t and decreasing in D_t. Note that our approach connects interest payments to the inverse of relative indebtedness, whereas only μ_D can be set by the government.

We require the representative agent's budget constraint to be of multiplicative form of disposable wealth in order to ensure a closed-form solution to her optimal consumption problem. Therefore, we rewrite the budget constraint by substituting Eq. (3.14) and Eq. (3.15) in Eq. (3.19) as

$$dW_t = ((1-\alpha)Y_t - D_t)\,(\mu_Y dt + \sigma_Y dB_t) + 2\mu_D D_t dt - c_t dt,$$

and restate

$$d\widetilde{W}_t = \widetilde{W}_t\,(\mu_Y dt + \sigma_Y dB_t) - \widetilde{c}_t dt, \quad \widetilde{W}_{t_0} := (1-\alpha)\,Y_{t_0} - D_{t_0}, \tag{3.20}$$

where $\widetilde{W}_t$ denotes the adjusted disposable wealth and $\widetilde{c}_t := c_t - 2\mu_D D_t$ adjusted consumption.[21]

We apply dynamic programming (see, e.g., Merton (1969)) to solve for the CRRA-

[21] Under standard as well as puttable debt, the necessary restatement of the corresponding budget constraints in Eq. (3.16) and Eq. (3.17) is achieved by setting $\widetilde{W}_{t_0} := (1-\alpha)Y_{t_0}$ and $\widetilde{c}_t := c_t - (\mu_D - R)D_t$. The resulting (restated) budget constraint is equal to Eq. (3.20).

representative agent's optimal consumption path

$$\widetilde{c}_t^\star = \arg\max_{\widetilde{c}_t} \mathbb{E}\left[\int_{t_0}^{t_0+1} u(\widetilde{c}_s, s)\mathrm{d}s + \bar{u}(\widetilde{W}_{t_0+1}, t_0+1)\right], \tag{3.21}$$

subject to the budget constraint in Eq. (3.20). It follows from CRRA-preferences that

$$u(\widetilde{c}_t, \widetilde{W}_t) = e^{-\rho t}\frac{\widetilde{c}_t^{1-\gamma}}{1-\gamma},$$

and utility of bequest

$$\bar{u}(\widetilde{W}_{t_0+1}, t_0+1) = e^{-\rho(t+1)} I(t_0+1)\frac{\widetilde{W}_{t_0+1}^{1-\gamma}}{1-\gamma},$$

for time preference $\rho \geq 0$ and risk aversion $\gamma > 0$, where $I(t_0+1)$ denotes the relative weight the consumer assigns to her end of period wealth.

Lemma 3. *The solution to the dynamic programming problem in Eq. (3.21) subject to the budget constraint in Eq. (3.20) is given by the optimal consumption plan*

$$\widetilde{c}_t^\star = \frac{\widetilde{W}_t}{\frac{1}{\phi} + e^{-\phi(t_0+1-t)}\left(I(t_0+1)^{\frac{1}{\gamma}} - \frac{1}{\phi}\right)}, \tag{3.22}$$

where

$$\phi = \frac{1}{\gamma}\left(\rho - (1-\gamma)\left(\mu_Y - \frac{\gamma\sigma_Y^2}{2}\right)\right).$$

Proof. For proof see Appendix 3.A. □

Finally, substituting $\widetilde{c}_t := c_t - 2\mu_D D_t$ into Eq. (3.22) yields the consumer's effective optimal consumption plan

$$c_t^\star = \frac{\widetilde{W}_t}{\frac{1}{\phi} + e^{-\phi(t_0+1-t)}\left(I(t_0+1)^{\frac{1}{\gamma}} - \frac{1}{\phi}\right)} + 2D_t\mu_D.$$

The variable c_t reflects the consumer's excess consumption and might take on negative values. In order to derive the utility of consumption, we add a base consumption level $\bar{c}$, assumed to be fixed in the short term (over one period). Adding $\bar{c}$ ensures that utility is derived from total consumption which itself is positive.

Second, having computed the optimal consumption path of the representative agent, we solve the government's optimization problem in terms of its optimal borrowing decision. The government aims to pick the optimal borrowing rate $\mu_D^\star$ that maximizes the expected utility of its risk-averse consumer while accounting for possible default costs.[22]

Once the government has chosen the optimal borrowing rate at t_0, the same cannot be changed until $t_0 + 1$. Hence, the government chooses $\mu_D^\star$ such that

$$\begin{aligned}\mu_D^\star &= \arg\max_{\mu_D} U_t\left(\mu_D\right)\\ &= \arg\max_{\mu_D} \mathbb{E}\left[\underbrace{\int_{t_0}^{t_0+1} u(c_s^\star + \bar{c}, s)\mathrm{d}s + \bar{u}(W_{t_0+1}, t_0+1)}_{\text{utility of consumption and bequest}} - \underbrace{e^{-\rho}\Psi\mathbf{1}_{\{\eta > D_{t_0}/Y_{t_0}\}}}_{\text{discounted costs of default}}\right], \qquad (3.23)\end{aligned}$$

where

$$\eta = \frac{D_{t_0+1} - (W_{t_0+1} - W_{t_0})}{Y_{t_0+1}}.$$

In Eq. (3.23), $c_s^\star$ denotes the representative agent's optimal time-dependent excess consumption at time s. W_{t_0+1} reflects the wealth at the end of the period as computed by $W_{t_0+1} = W_{t_0} + \int_{t_0}^{t_0+1} \mathrm{d}\widetilde{W}_s$. The last term in Eq. (3.23) corresponds to a risk-neutral government's expected default costs discounted to time t_0, imposing equal time preferences as for the representative consumer.

Proposition 2. *If expected default costs under GDR bonds are sufficiently high to incentivize the government to choose a bounded $\mu_D^\star$, then this $\mu_D^\star$ is unique.*

Proof. For proof see Appendix 3.A. □

In summary, when choosing its optimal borrowing rate, a heavily indebted government needs to optimally trade off higher accumulated consumption versus an increasing probability of default. GDR bonds might provide more space to reduce sovereign borrowing, while not too severely limiting consumption. Therefore, GDR bonds could potentially allow for more sustainable deleveraging policies.

3.3.3 Comparative Statics

The comparative statics analysis of the above refinancing instruments requires an analogous result as in Proposition 2 but for standard and puttable debt.

Corollary 1. *If expected default costs under standard debt and puttable debt are sufficiently high to incentivize the government to choose a bounded $\mu_D^\star$, then this $\mu_D^\star$ is unique.*

Proof. For proof see Appendix 3.A. □

[22] Similar to Müller et al. (2016), we also model welfare to be additively separable in the utility of consumption and a linear default costs component.

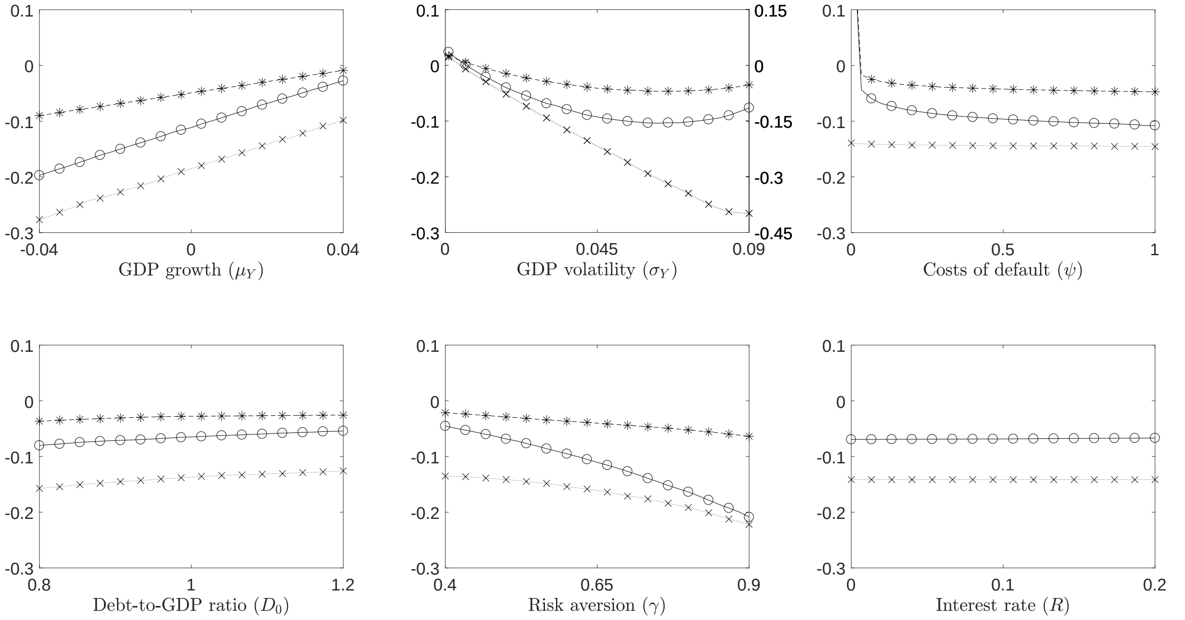


Figure 3.3: Comparative statics of the optimal borrowing rate

Notes: This figure shows the government's optimal borrowing rate $\mu_D^\star$ under the three refinancing instruments, i.e., standard sovereign debt (solid, circles), puttable debt (dotted, crosses), and GDR bonds (dashed, asterisks), with respect to model parameters. Initial values: $\mu_Y = 0.02$, $\sigma_Y = 0.03$, $Y_0 = 1$, $D_0 = 0.95$, $\alpha = 0.03$, $\rho = 0.03$, $\gamma = 0.5$, $I(t_0 + 1) = 10$, $\Psi = 0.1$. Note that, for GDP volatility, $\mu_D^\star$ under puttable bonds is depicted on the right axis.

Figure 3.3 presents the comparative statics of a government's optimal borrowing rate $\mu_D^\star$ under standard sovereign debt, puttable debt, and GDR bonds.[23] Similarly, Figure 3.4 displays the comparative statics of the corresponding default probabilities given the government's optimal borrowing rate $\mu_D^\star$. By referring to the various plots in Figure 3.3 and Figure 3.4, we are able to make several interesting statements regarding the sensitivity of $\mu_D^\star$ and the implied default probability with respect to changes of model parameters:

1. Unsurprisingly, the optimal borrowing rate is increasing in GDP growth. This is intuitive since a higher GDP growth rate allows to accelerate debt-financed consumption without increasing relative indebtedness. Thanks to lower interest expenses under low growth, relying on GDR bonds enables the government to borrow

[23] Theoretically, by applying the implicit function theorem, one can derive the sensitivity of the optimal borrowing rate with respect to model parameters. However, since the optimal borrowing rate depends on intertemporal utility of consumption, we cannot compute closed-form solutions as, to the best of our knowledge, its integral over time can only be solved numerically. This prevents us from applying the implicit function theorem with respect to most parameters. There is one important exception: We can compute the sensitivity of the optimal borrowing rate with respect to default costs under all three refinancing instruments, as they only influence the expected default costs but not intertemporal consumption. We find that $\partial\mu_D^\star/\partial\psi < 0 \; \forall \psi \in \mathbb{R}^+$. Derivations are available on request.

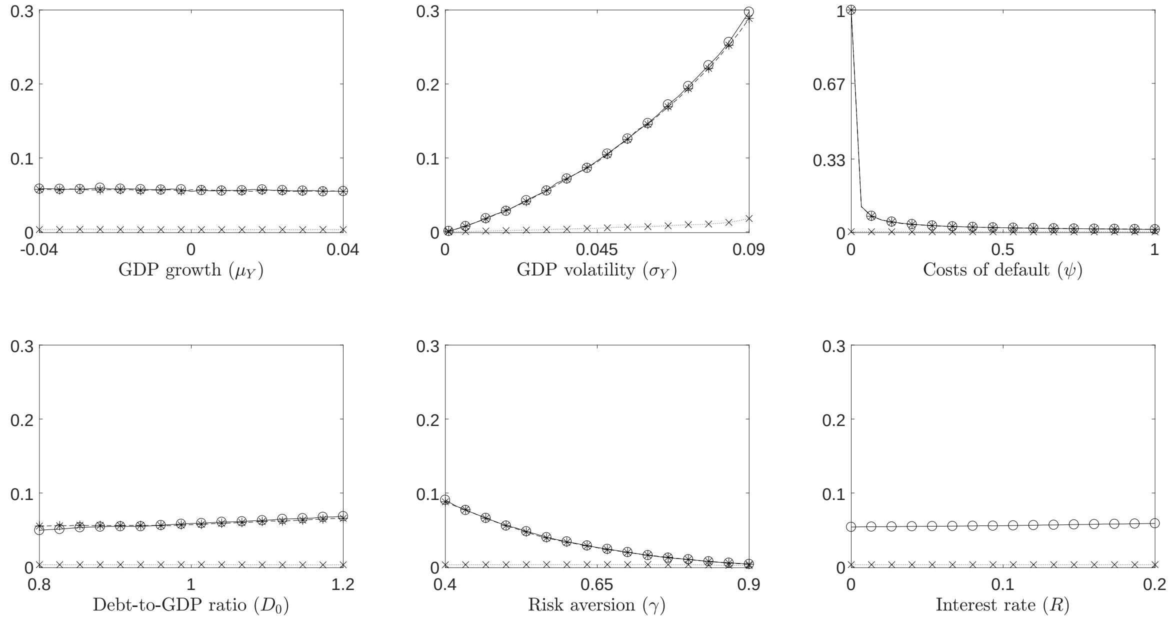


Figure 3.4: Comparative statics of the probability of default given optimal borrowing

Notes: This figure shows the government's probability of default given given the optimal borrowing rate $\mu_D^\star$ under the three refinancing instruments, i.e., standard sovereign debt (solid, circles), puttable debt (dotted, crosses), and GDR bonds (dashed, asterisks), with respect to model parameters. Initial values: $\mu_Y = 0.02$, $\sigma_Y = 0.03$, $Y_0 = 1$, $D_0 = 0.95$, $\alpha = 0.03$, $\rho = 0.03$, $\gamma = 0.5$, $I(t_0+1) = 10$, $\Psi = 0.1$.

more than under standard or puttable debt without increasing the implied default probability. Regardless of the level of μ_Y, the probability of default remains very low under standard debt and GDR bonds, and close to zero under puttable debt. A change in the GDP growth rate has a negligible impact on the default probability under all three refinancing instruments.

2. The uncertainty of future GDP growth and the optimal borrowing rate under standard and GDR bonds feature a U-shaped relationship. For very low levels of σ_Y, i.e., in an almost deterministic model, we observe an optimal borrowing rate that induces the debt-to-GDP ratio to approach but not to cross the default threshold. However, once the growth shocks become non-negligible, the government's optimal borrowing rate decreases significantly. In order to prevent the risk of large enough negative shocks that could trigger default, it is optimal to preemptively smooth out the potential destabilizing effects on relative indebtedness by borrowing less. However, GDR bonds' variable interest rates alleviate the risk of such shocks, thereby lowering the government's need to further reduce borrowing.

   Once GDP volatility exceeds a certain level, the risk that the country ends up in default starts to increase significantly. If the government wanted to further limit the probability of default, massive and costly (in terms of foregone consumption) cuts in the optimal borrowing rate were needed. As a result, the government's optimal borrowing level starts increasing again when the chance to enter default becomes substantial (i.e., more than $> 20\%$ for the given parameters). Under puttable debt, the optimal borrowing rate is decreasing in growth volatility since higher uncertainty leads to a higher payable put price. As a result, the government chooses a very low borrowing rate in order to sustain a tiny default probability, thereby minimizing its insurance costs.

3. Regardless of the refinancing instrument, the implied optimal borrowing rate is decreasing in the imposed default costs. The higher these costs, the higher are the incentives not to enter default. As a consequence, the government borrows less if the costs of default are high. Similarly, the probability of default is decreasing in the associated costs, hence the country minimizes its default risk once the corresponding costs are substantial enough.

4. The initial debt-to-GDP ratio's effect on optimal borrowing is slightly positive. The higher this ratio, the smaller is the impact of a negative growth shock. As a consequence, the government can marginally increase borrowing as the default risk becomes less sensitive to growth uncertainty. In addition, the marginal return on borrowing is increasing in the initial debt level, which motivates, ceteris paribus, higher borrowing costs. Both effects seem counterproductive, as countries with higher indebtedness borrow more. However, the first effect is only due to our simplified definition of sovereign default. Moreover, the optimal borrowing rate is negative for all levels of the initial debt-to-GDP ratio. The relative indebtedness is hence even decreasing for high levels of debt, just at a slower pace. The effect of the initial debt-to-GDP ratio on the probability of default is tiny.

5. The consumer's risk aversion inversely affects the optimal borrowing rate. If less risk averse, she prefers higher (debt-financed) consumption today at the cost of a higher default probability tomorrow. For a strongly risk-averse consumer, by contrast, the government decreases its borrowing rate in order to reduce the implied default risk. Puttable debt's inherent insurance mechanism as well as GDR bonds' interst rate structure both alleviate the optimal borrowing rate's sensitivity to the consumer's risk aversion.

6. Under standard sovereign debt, the optimal borrowing rate is slightly increasing in the interest rate. The higher the interest payments, the more does the government

need to borrow in order to maintain consumption. Under puttable debt, the government pays the risk-free rate on its bonds. Hence, R only affects the value of the put as we assume that the government has to compensate the insurance guarantor for its initial default risk (see Section 3.2), thereby conservatively neglecting a lower interest rate's decreasing effect on the ex-post default probability. However, the effect of the interest rate on the government's optimal borrowing rate under puttable debt is negligible.[24]

In summary, we find two persistent patterns across all parameter variations. First, optimal borrowing is highest under GDR and lowest under puttable bonds. GDR bonds allow the government to borrow more without substantially increasing its default probability compared to standard debt, thereby making the deleveraging committment more feasible. Under puttable debt, by contrast, the government extends its deleveraging effort in order to reduce the imposed insurance costs. Second, we observe the default probability to be almost equal for standard and GDR bonds but substantially lower for puttable bonds. This indicates that the latter's ex-ante payable insurance premium incentivizes the government to choose very low borrowing rates which in turn almost completely eliminate its default risk. Specifically, the reduction in borrowing further amplifies the default probability decreasing effect of puttable bonds' per se lower interest rates.

3.4 Case Study: Portugal, Ireland, Italy, Greece, and Spain

We now investigate GDR bonds' potential virtue of mitigating the risk of swelling refinancing costs in the presence of high indebtedness by analyzing the cases of Portugal, Ireland, Italy, Greece and Spain at the time of the OMT announcement on July 26, 2012. We base our analysis on a comparison of the model predicted effectiveness of the following three refinancing instruments: (i) the status quo with standard debt, (ii) puttable (risk-free) debt in return for an ex-ante insurance premium in the spirit of Section 3.2, and (iii) GDR bonds as introduced in Section 3.3. Like in Section 3.3, we analyze those three refinancing instruments in an environment with continuous consumption and existing default costs.

In 2010, under the status quo, highly indebted Eurozone countries raised concerns regarding sovereign default after having reportedly been targeted by speculative attacks

[24] Similarly, the effect of the consumer's time-preference is rather weak and only plays a minor role in determining $\mu_D^\star$ and the corresponding default probability. Marginally, the lower the consumer's patience, i.e., the higher ρ, the more she wants to consume today, leaving the government to prefer a higher optimal borrowing rate.

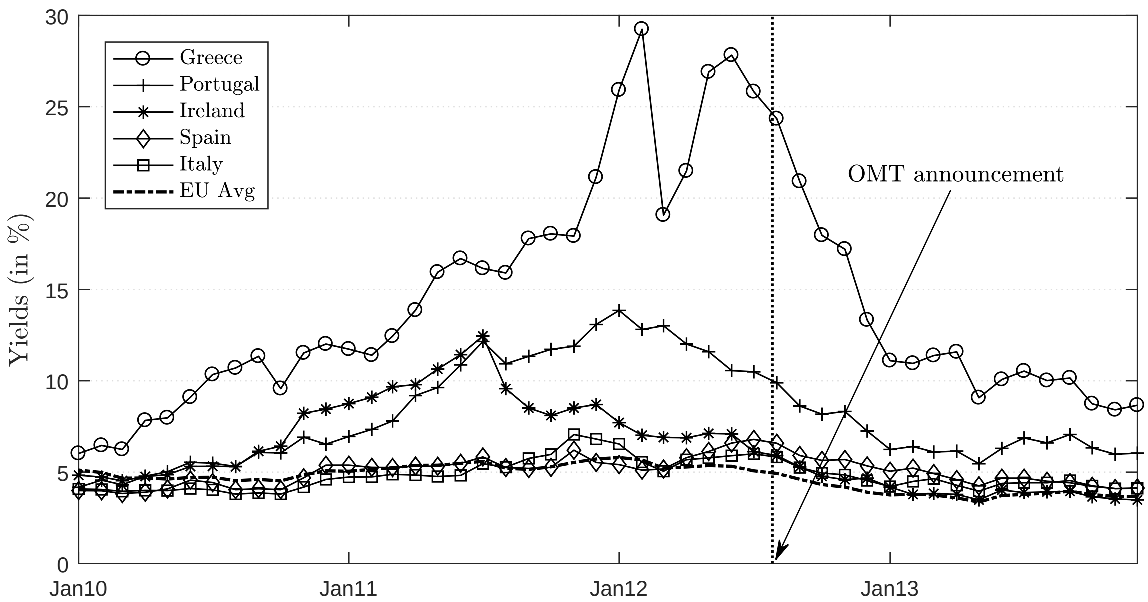


Figure 3.5: 10-year yields on selected government bonds

Notes: This figure shows 10-year yields on government bonds for Portugal, Ireland, Italy, Greece, and Spain based on ECB data.

on their bond yields.[25] Such attacks, that ultimately resulted in an upwards shift of the country's perceived default risk, lead to higher demanded yields, which ultimately increase—likely amplified by the currency union's internal flight-to-safety—its actual default probability. Indeed, the empirical evidence of De Grauwe and Ji (2012) suggests a significant effect of negative market sentiments on the spreads of Greece, Ireland, Portugal, and Spain. In particular, the sovereign bond yields of Portugal, Ireland, and Greece soared to unprecedented levels due to fears of imminent sovereign default. By the end of 2011, as shown in Figure 3.5, long-term bond yields had climbed to 13.08% for Portugal, 8.70% for Ireland, and even 21.14% for Greece, respectively.[26] At these levels, refinancing became very expensive, which finally led the ECB to announce its OMT program in the summer of 2012.

Table 3.3 provides an overview of the calibrated parameters for the five countries.[27]

[25] As highlighted by both the IMF (source: IMF Global Financial Stability Report, April 2012, http://www.imf.org/external/pubs/ft/gfsr/2012/01/) and the ECB (source: Draghi, M., ECB press conference, September 6, 2012, http://www.ecb.europa.eu/press/pressconf/2012/html/is120906.en.html).

[26] Despite the rise in yields, the governments of Portugal, Ireland, Italy, Greece, and Spain increased sovereign borrowing over the same time period, indicating that higher refinancing costs force countries to borrow more in order to finance consumption (source: Eurostat, http://ec.europa.eu/eurostat/en/web/government-finance-statistics/statistics-illustrated).

[27] Our results are robust to changes in the standard values of parameters ρ, γ, and $I(t_0+1)$.

Since there is a monotone one-to-one relation between a country's optimal borrowing rate and its default costs (see Footnote 23), we calibrate the latter such that its respective default probability under the status quo is equal to one implied by its actual sovereign credit rating prior to the OMT announcement. We therefore borrow Moody's historical default probabilities, implying a default probability of 1.9% for a Baa rating (Italy and Spain), 6.3% for a Ba rating (Portugal and Ireland), and 33.3% for a C rating (Greece), respectively.[28] Unfortunately, historical default probabilities are not available for subgroups. For instance, we cannot distinguish between the default probabilities of Portugal (Ba3) and Ireland (Ba1). Clearly, Portugal's higher interest rate (10.2% versus 5.2%) indicates a higher default probability.[29] Except for Greece, we use 4-year generic government bond yields in order to be in line with Collard et al. (2015), who rely on an average duration of sovereign debt equal to four years. For Greece, due to limited data availability, we rely on its 10-year government bond yield.

We set $I(t_0+1) = 7$, as we find this level to provide sufficient incentives to save a significant part of disposable wealth for future periods. Given this weight on the consumer's end of period wealth, the expected disposable wealth is slightly increasing (on average 0.04% under the status quo). In addition, a comparative statics analysis with respect to $I(t_0+1)$ reveals that the sensitivity of the optimal borrowing rate to changes in the former is very small and negligible for levels of $I(t_0+1)$ around seven. This is desirable as we do not want this parameter choice to influence our results with respect to the optimal borrowing policies.

In addition, under (iii) GDR bonds, we need to ensure that the initial value $\widetilde{W}_{t_0}$ is nonnegative. A negative initial value of the process $\widetilde{W}_t$ would imply that higher GDP growth and more borrowing decrease disposable wealth for consumption. As a consequence, we limit GDR debt issuance to the maximum value such that $\widetilde{W}_{t_0}$ is nonnegative.

In order to test the three different sovereign borrowing policies for different GDP growth realizations, we introduce a scenario analysis with respect to the realizations of μ_Y. We consider three realizations of the GDP growth rate: (i) the best out of the 5% worst realizations of μ_Y, i.e., the fifth historical percentile,[30] (ii) zero-growth, and (iii) the historical average growth rate. We then compare overall utility of the three refinancing instruments under these three scenarios. Figure 3.6 displays the resulting differences in utility for Portugal, Ireland, Italy, Greece, and Spain.

---

[28] Source: Moody's, Sovereign Default and Recovery Rates, 1983-2008, March 2009, https://www.moodys.com/sites/products/DefaultResearch/2007400000587968.pdf.

[29] We additionally analyze the three refinancing instruments' effectiveness for probabilities of default as implied by traded CDS-spreads and find similar results.

[30] We again assume a normally distributed growth rate, implying the fifth percentile to equal $\mu_Y - 1.645\sigma_Y$, where μ_Y and σ_Y denote the historical average GDP growth rate and volatility based on yearly OECD data (http://stats.oecd.org).

Table 3.3: Parameters

Parameter	Variable	Portugal	Ireland	Italy	Greece	Spain
GDP growth	μ_Y	2.29%	3.35%	1.68%	1.57%	2.01%
GDP volatility	σ_Y	3.46%	3.50%	2.31%	3.68%	2.23%
Debt-to-GDP ratio	D_0/Y_0	102.41%	110.55%	108.35%	111.11%	61.83%
Interest rate	R	10.22%	5.19%	5.85%	27.82%	6.95%
MPS	α	0.23%	6.74%	6.51%	4.37%	4.01%
Consumption	$\bar{c}$	65.82%	45.57%	61.51%	69.88%	57.82%
Risk aversion	γ	0.5	0.5	0.5	0.5	0.5
Time preference	ρ	5.00%	5.00%	5.00%	5.00%	5.00%
Default costs	Ψ	0.13	0.21	0.42	0.02	0.28
Risk free rate	r	1.30%	1.30%	1.30%	1.30%	1.30%

Notes: This table presents the calibrated parameters for Portugal, Ireland, Italy, Greece, and Spain. We use yearly OECD (http://stats.oecd.org) GDP data for the years 1970-2011 to compute μ_Y and σ_Y. D_0/Y_0 corresponds to relative indebtedness as of the end of 2011 and is based on World Bank data (http://data.worldbank.org). We use yield data of 4-year generic government bonds per July 26, 2012 from Bloomberg for all countries except Greece. As there is no data on traded government bonds with maturity less than ten years for Greece at that time, we use 10-year government bond yields provided by the Bank of Greece (http://www.bankofgreece.gr). MPS corresponds to the maximum primary surplus based on OECD data. Consumption denotes national consumption in the year 2011 based on World Bank data. Standard values are assumed for γ and ρ. Default costs are set such that the default probability under the status quo and historical GDP growth matches the sovereign default probability according to Moody's credit rating at the time of OMT announcement. The continuously compounded yield on a German 4-year government bond as of July 26, 2012 serves as a proxy for the risk-free interest rate r.

We find that, across growth scenarios, switching from standard to puttable debt leads to welfare improvements. The savings from lower interest payments more than compensate the government for its ex-ante insurance payment. Due to countries' low borrowing rates, the value of the put option is relatively low considering prevailing interest rate levels, ranging from 0.18% (Spain) to 0.76% (Ireland) of GDP. The effect of the three different GDP growth scenarios on the insurance premium is small. In addition, default probabilities under puttable debt (0.48% on average) are substantially lower than under the status quo (11.05%). This finding is consistent with the result from Section 3.2, i.e., that, under the presence of default costs, puttable debt's default risk mitigation is welfare improving. Only Ireland, due to its comparably low interest rate, does not experience a substantial utility increase when issuing puttable bonds.

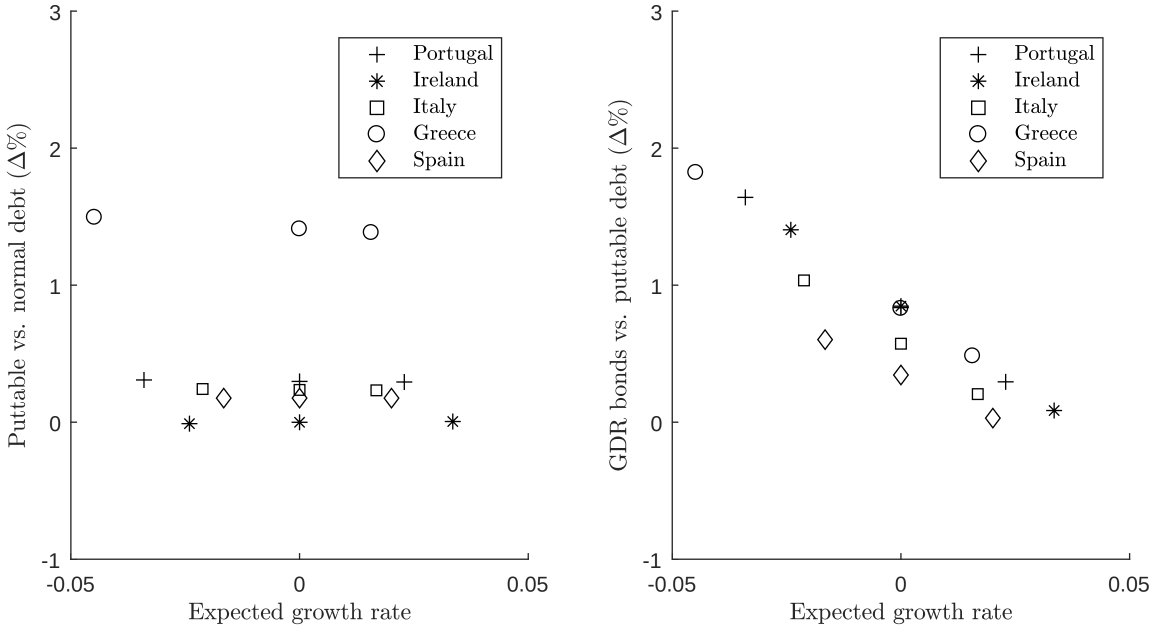


Figure 3.6: Utility comparison between puttable and normal bonds (left) as well as GDR and puttable bonds (right)

Notes: The left-hand side plot provides the relative utility differences between puttable and standard debt. The right-hand side plot displays relative utility differences between GDR bonds and puttable debt.

Moreover, Figure 3.6 highlights that GDR bonds consistently outperform puttable bonds. The welfare improvements induced by the former are especially pronounced for the case of zero or negative GDP growth, where GDR bonds imply significantly lower interest payments. Thus, contrary to puttable debt's simple insurance mechanism, GDR bonds' state-contingent interest charges allow the risk-averse consumer to considerably smooth her within-period consumption path, without increasing associated default risks too heavily. On average, optimal borrowing rates are approximately twice as low under standard debt than under GDR bonds, and even two times lower for puttable relative to standard debt. In summary, the data-implied default costs are not high enough such that puttable debt's reduction in default risk would keep up with GDR bonds' capability to smooth consumption.

If one instead considers more pessimistic (risk-neutral) default probabilities implied by historical CDS spreads, the price of puttable debt's embedded put option becomes larger. This can change the relation between puttable and standard debt: As the ex-ante payable insurance premium increases, wealth disposable for consumption is reduced, making standard debt relatively more attractive. However, GDR bonds are still outperforming both normal and puttable bonds, leaving the above results robust to changes in the calibration

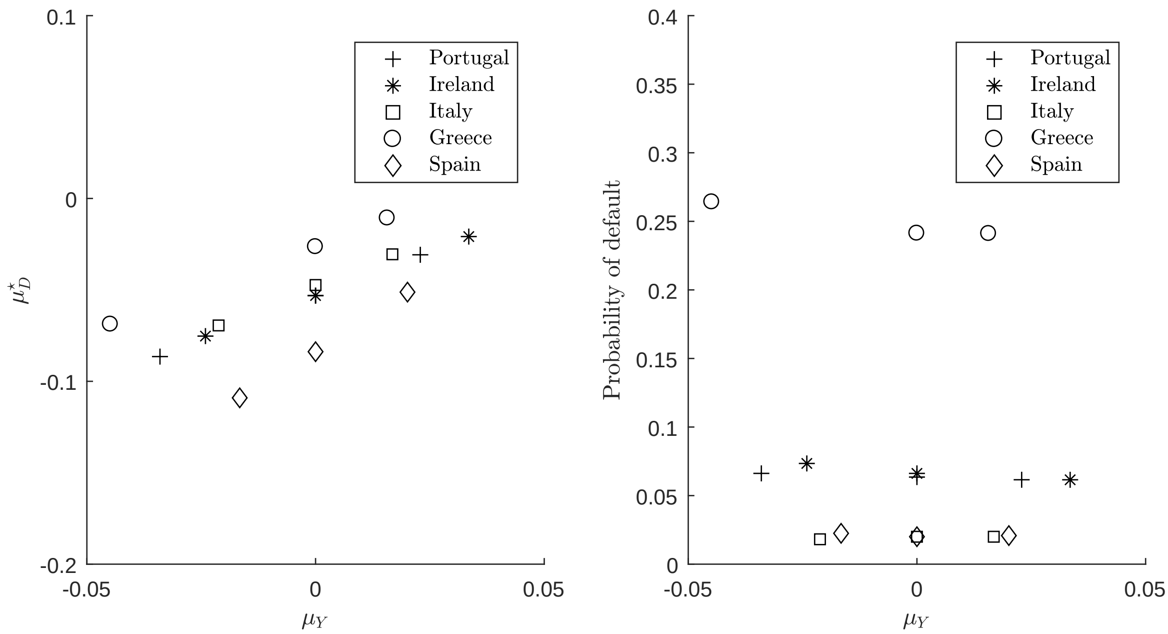


Figure 3.7: Optimal borrowing rate (left) and default probability (right) under GDR bonds

Notes: The left-hand side plot shows the optimal borrowing rate of Portugal, Ireland, Italy, Greece, and Spain under GDR bonds. The right-hand side plot displays the respective probabilities of sovereign default.

of default probabilities.[31]

To sum up, it hardly depends on the realization of the GDP growth rates which instrument is most beneficial. We find the status quo with standard debt to be inferior to puttable debt. Furthermore, GDR bonds which are inversely linked to relative indebtedness, provide even higher utility than the issuance of puttable debt. Retrospectively, issuing GDR bonds might have been an expedient refinancing alternative for the herein considered countries.

Having selected our preferred refinancing instrument, we plot the optimal borrowing rates under GDR bonds as well as the corresponding default probabilities for the three GDP growth scenarios in Figure 3.7. We find the optimal borrowing rates to lie between -10.90% (Spain with $\mu_Y = -1.66\%$) and -1.07% (Greece with $\mu_Y = 1.57\%$). Hence, GDR bonds' incentive structure appears sufficient in inducing the issuing government to

[31] On average, imposing CDS-implied default probabilities, i.e., lowering Ψ, increases utility under normal and GDR bonds by 0.13% and under puttable bonds by 0.01%. As a consequence, puttable bonds are relatively less attractive under CDS-implied default probabilities. This discrepancy to Section 3.2 arises due to the significantly lower default costs and the high weight on the consumer's reduced utility of bequest, which is absent in Section 3.2's analysis of puttable debt. More importantly, under lower values of Ψ, GDR bonds still allow for higher utility levels relative to normal and puttable debt.

Table 3.4: Sharpe ratios under GDR bonds: Scenario analysis

	Negative GDP growth	Zero GDP growth	Positive GDP growth
Portugal	1.48	1.50	1.51
Ireland	1.42	1.48	1.51
Italy	2.03	1.99	1.99
Greece	0.61	0.68	0.68
Spain	4.09	3.70	3.14

Notes: This table displays the Sharpe ratios of GDR bonds as computed by $SR = (\mu_Y - \mu_D^\star - r_f)/\sigma_Y$, where $\mu_D^\star$ reflects the optimal borrowing rate of the government. The negative GDP growth scenario features the best out of the 5% worst historical realizations of μ_Y, i.e., the fifth historical percentile, zero GDP growth implies $\mu_Y = 0$, and positive GDP growth corresponds to the historical average growth rate. We use yearly GDP growth and volatility data from the OECD (http://stats.oecd.org).

choose a deleveraging borrowing rate. Figure 3.7 shows that neither the absolute level of the probability of default nor the order of the five countries depend on the GDP growth rate.

Finally, we compute the Sharpe ratio of GDR bonds in order to evaluate their potential attractiveness to risk-averse investors. The Sharpe ratios equal $(\mu_Y - \mu_D^\star - r_f)/\sigma_Y$, where $\mu_Y - \mu_D^\star$ reflects GDR bonds' expected interest rate. If the Sharpe ratio of such an investment instrument was below a minimum threshold, investors would not be willing to provide funds, as their expected compensation would be too small given the associated risk level. Table 3.4 presents the computed Sharpe ratios for the three growth scenarios considered above.

Overall, the Sharpe ratios are high, i.e., between 0.61 and 4.09. Regardless of the realization of the GDP growth rate, GDR bonds appear to be an attractive investment instrument, since the optimal borrowing rates are always negative. In sharp contrast to standard GDP-indexed bonds, GDR bonds' Sharpe ratios remain competitive even for negative growth rates.

However, if governments possibly deviated from their initial optimal borrowing rate *after* having issued GDR bonds, their actual risk-return profile could deteriorate.[32] If markets deem this risk unacceptable, establishing an intergovernmental agency as counterparty (writer) of GDR embedded put options could foster the credibility of initially proclaimed deleveraging intentions. Buying such a put option would give hesitant investors the right to hand over their GDR bonds to the option writer for a fixed amount, in

[32] This, in fact, could happen if Ψ decreases or if default costs are not enforceable (see discussion in Section 3.5).

case prices fell sharply due to a too high μ_D. Hence, the put option would insure investors against overborrowing by the GDR bonds issuing government.

In addition, the put premium would fairly compensate the option writing counterparty for guaranteeing credit insurance. Importantly, the latter nevertheless had to maintain the enforceability of Ψ upon option exercise. The case of Greece provides indicative evidence that such intergovernmental institutions, i.e., the European Commission, Eurogroup, ECB, and IMF, indeed possess enough power to conditionally force Eurozone debtor countries to massive cuts in fiscal spending. In Section 3.5 below, potential risks associated with the implementation of GDR bonds are addressed more generally.

3.5 Discussion and Concluding Remarks

In summary, our analysis indicates considerable virtues of GDR over puttable bonds in terms of welfare improvements. Naturally, the introduction of a novel financial contract such as GDR bonds brings about many new factors, which all need to be considered in detail. In this final section, we address the most apparent points. We argue that—albeit justified—they can be dealt with in reasonable manner.

Risk of Limited Commitment

Given the inverse relation between interests paid on GDR bonds and a country's current GDP-to-debt ratio, a natural concern is that issuing governments simply keep on increasing their debt to lower payable interests on outstanding GDRs. From an ex-ante perspective, this is no problem since investors are only willing to purchase GDRs, if the government can credibly commit itself to a deleveraging borrowing rate. Ex-post, however, a limited commitment problem due to a lack of enforceability may arise.

The enforceability of non-positive borrowing rates could, e.g., be achieved by contractually forbidding the issuing government to auction any non-authorized debt during the GDRs' lifetime. For this purpose, an intergovernmental organization would have to be established, similar to, e.g., the former European Troika (recently renamed as 'European Quadriga'), who enforces compliance with such contractual agreements. To be itself credible, it needs to be empowered to impose sanctions, e.g., predefined coercive measures, in case of contravention.

Alternatively, similar to puttable debt, GDR bonds could be complemented by an embedded put option. Whenever the government's GDP-to-debt ratio at GDRs' maturity falls below a certain threshold, investors could put their bonds with the above mentioned organization and receive a fixed payment in return. In case their GDR bonds get put, sanctions are imposed on the issuing government. Note that, contrary to puttable debt discussed above, the writer of the embedded put option would be compensated with the insurance premium paid by investors instead of the issuing government.

Growth Risk

By construction, returns to GDR bonds are pro-cyclical, i.e., if growth rates are positive (negative), interests increase (decrease). Hence, similar to GDP-linked bonds (see Blanchard et al. (2016)), GDRs represent so-called 'high beta' instruments, for which investors demand higher expected returns per unit of risk.

In light of the above calibrated Sharpe ratios, we argue that GDRs easily exceed the necessary return-for-risk profile usually demanded by investors for 'high beta' securities. Particularly, GDRs' Sharpe ratios are generally high due to the relatively low levels of growth volatility. Moreover, growth rates are only imperfectly correlated across countries. Hence, making GDRs available to a widely dispersed foreign investor base further reduces country-specific growth risks borne by investors.

Novelty and Liquidity Risk

Newly introduced GDR bonds would entail substantial novelty and liquidity risks. Historically, GDP-linked bonds have been mainly issued by developing countries, often during debt restructuring. In the context of the potential isssuers we have in mind, Blanchard et al. (2016) conclude: "With relatively strong institutions and an independent statistical agency, advanced economies are in a better position to give confidence to investors that data on economic growth will remain untampered and reliable." (Blanchard et al., 2016, p.3)

The absence of a reasonably liquid secondary market usually complicates the introduction of novel financial securities, such as GDP-linked bonds (Barr et al. (2014)). A minimum scale and sufficient standardization would therefore facilitate the successful implementation of GDRs. The Sharpe ratios calibrated for Portugal, Ireland, Italy, Greece, and Spain firmly appear high enough to compensate for GDRs' potential novelty and liquidity risks. In case of unexpectedly high initial liquidity concerns and novelty aversion, making GDRs puttable, as discussed above, jointly decreases both types of risks.

In conclusion, we believe GDP-to-debt-index bonds to constitute a promising alternative for managing the risk of future spikes in government yields for highly indebted Eurozone countries, of whom there still are many. If implemented appropriately, they could be an effective state-contingent debt instrument, temporary mitigating refinancing pressure and smoothing private consumption, thereby creating space and time for a less painful, but more sustainable sovereign deleveraging.

3.A Appendix: Proofs

Proof of Lemma 1. We start from Eq. (3.12), rearranging yields

$$\beta_M \left(\frac{\pi_L}{\alpha + \frac{\beta_M}{1+rL}} + \frac{\pi_H}{\alpha + \frac{\beta_M}{1+rH}} \right) = \frac{(1+\mu_g+\sigma_g)^2}{8\sigma_g} \quad \Leftrightarrow$$

$$\frac{\beta_M^2(\mathbb{E}[1+r]) + \beta_M(\alpha(1+r_L)(1+r_H))}{\beta_M^2 + \beta_M\alpha(2+r_L+r_H) + \alpha^2(1+r_L)(1+r_H)} = \frac{(1+\mu_g+\sigma_g)^2}{8\sigma_g},$$

which leads us to the following quadratic equation

$$\beta_M^2 \underbrace{\left(\frac{(1+\mu_g+\sigma_g)^2}{8\sigma_g} - \mathbb{E}[1+r] \right)}_{=:A} + \beta_M \underbrace{\left(\frac{(1+\mu_g+\sigma_g)^2}{8\sigma_g} \alpha(2+r_L+r_H) - \alpha(1+r_L)(1+r_H) \right)}_{=:B}$$
$$+ \underbrace{\frac{(1+\mu_g+\sigma_g)^2}{8\sigma_g} \alpha^2(1+r_L)(1+r_H)}_{=:C} = 0. \tag{3.A.1}$$

Whenever $A \geq 0$, it must be that

$$\frac{(1+\mu_g+\sigma_g)^2}{8\sigma_g} \geq \mathbb{E}[1+r] \geq 1,$$

where the second inequality holds since $0 \leq r < 1$. Hence, $A \geq 0$ directly implies $B > 0$. However, if both A and B are nonnegative, there exists no nonnegative solution β_M to Eq. (3.A.1). A nonnegative β_M therefore requires that $A < 0$, i.e.,

$$\mathbb{E}[1+r] > \frac{(1+\mu_g+\sigma_g)^2}{8\sigma_g},$$

and because $C \geq 0$, the maximum β_M is given by

$$\beta_M = \frac{-B - \sqrt{B^2 - 4AC}}{2A} > 0.$$

This completes the proof. □

Proof of Lemma 2. We prove Lemma 2 in two steps. First, we compute the date-t put option value. A standard European put option with strike price x, underlying S, and maturity $t+1$ yields the following payoff at expiration

$$\max(x - S_{t+1}, 0).$$

In the context of puttable bonds, this payoff can be written as

$$\max\left(b_M(t)(1+R_H)Y_t - \alpha Y_{t+1}, 0\right).$$

In the following, we denote by $\underline{g}$ the growth-rate threshold below which the put option yields a positive payoff at maturity, i.e., the minimum growth rate necessary to prevent government default:

$$b_M(t)(1+R_H)Y_t = (\alpha + b_M(t+1))Y_t\underline{g} \quad \Leftrightarrow \quad \frac{b_M(t)(1+R_H)}{\alpha + b_M(t+1)} = \underline{g}(r_{t+1}).$$

Since $\underline{g}$ is state-dependent, we write it as a function of r_{t+1}. Given risk-neutrality and that $r_t = r_H$, the date-t put price then equals

$$\begin{aligned}
&\delta\left(\mathbb{P}\left(g < \underline{g}(r_{t+1})\right) \times E_t\left[b_M(t)(1+R_H)Y_t - \alpha Y_{t+1} | g < \underline{g}(r_{t+1})\right]\right) = \\
&\frac{1}{1+r_H}\left(\pi_L \frac{\underline{g}(r_L) - 1 - \mu_g + \sigma_g}{2\sigma_g} \times \left(b_M(t)(1+R_H) - \alpha\frac{1+\mu_g - \sigma_g + \underline{g}(r_L)}{2}\right) Y_t + \right. \\
&\qquad \left. \pi_H \frac{\underline{g}(r_H) - 1 - \mu_g + \sigma_g}{2\sigma_g} \times \left(b_M(t)(1+R_H) - \alpha\frac{1+\mu_g - \sigma_g + \underline{g}(r_H)}{2}\right) Y_t\right),
\end{aligned} \tag{3.A.2}$$

where the conditional expectation can be calculated in closed form by relying on the uniform distribution of g. Recalling the assumption of zero recovery in default from Section 3.2, Eq. (3.A.2) finally implies

$$p_t = \frac{1}{1+r_H}\left(\pi_L \frac{\underline{g}(r_L) - 1 - \mu_g + \sigma_g}{2\sigma_g} + \pi_H \frac{\underline{g}(r_H) - 1 - \mu_g + \sigma_g}{2\sigma_g}\right) \times b_M(t)(1+R_H)Y_t. \tag{3.A.3}$$

Second, we show that the government is indeed indifferent between issuing normal and puttable debt, i.e., that the following equation holds:

$$p_t = \underbrace{\delta}_{=\frac{1}{1+r_H}} b_M(t)(R_H - r_H)Y_t. \tag{3.A.4}$$

Plugging Eq. (3.A.3) into Eq. (3.A.4) and after some tedious but simple algebra wet get

$$\underbrace{b(t)(1+r_H)}_{\equiv \beta_M} = \left(8\sigma\left(\frac{\pi_L}{\alpha + \frac{\beta_M}{1+r_L}} + \frac{\pi_H}{\alpha + \frac{\beta_M}{1+r_H}}\right)\right)^{-1} \frac{1+r_H}{1+R_H}\left(\frac{2\sigma(R_H - r_h)}{1+R_H} + 1 + \mu_g - \sigma_g\right). \tag{3.A.5}$$

Now, Eq. (3.12) implies

$$R_H := \frac{d_M}{b_M(t)} - 1 = \frac{d_M}{\beta_M/(1+r_H)} - 1 = \frac{4\sigma_g(1+r_H)}{1+\mu_g+\sigma_g} - 1. \tag{3.A.6}$$

Plugging Eq. (3.A.6) into Eq. (3.A.5) and simplifying finally yields

$$\beta_M = \frac{(1+\mu_g+\sigma_g)^2}{8\sigma\left(\frac{\pi_L}{\alpha+\frac{\beta_M}{1+r_L}} + \frac{\pi_H}{\alpha+\frac{\beta_M}{1+r_H}}\right)}.$$

This completes the proof. □

Proof of Proposition 1. In the absence of puttable bonds, if $r_t = r_H$, the government defaults at $t+1$ whenever

$$b_M(t)(1+R_H)Y_t > (\alpha + b_M(t+1))Y_t g_{t+1} \quad \Leftrightarrow \quad \frac{b_M(t)(1+R_H)}{\alpha+b_M(t+1)} > g_{t+1},$$

i.e., the probability of default is given by

$$\mathbb{P}|R_H := \mathbb{P}\left(g < \frac{b_M(t)(1+R_H)}{\alpha+b_M(t+1)}\right).$$

Given the uniform distribution of g, we get

$$\mathbb{P}|R_H = \frac{\frac{b_M(t)(1+R_H)}{\alpha+b_M(t+1)} - (1+\mu_g-\sigma_g)}{2\sigma_g}. \tag{3.A.7}$$

We have to distinguish between two cases: First, if $b_M(t)(1+r_H)/(\alpha+b_M(t+1)) > 1+\mu_g-\sigma_g$, Eq. (3.A.7) implies that issuing puttable instead of standard debt reduces the default probability by

$$\mathbb{P}|R_H - \mathbb{P}|r_H = \frac{b_M(t)(R_H - r_H)}{2\sigma_g\left(\alpha+b_M(t+1)\right)}.$$

Second, in the complementary case, i.e., if $\mathbb{P}|r_H = 0$, puttable bonds eliminate the whole default risk imposed by standard bonds. This completes the proof. □

Proof of Lemma 3. In order to solve for optimal consumption, we apply standard dynamic programing techniques as described in, e.g., Merton (1969). We want to solve for

$$\widetilde{c}_t^\star = \arg\max_{\widetilde{c}_t} U(\widetilde{c}_t, \widetilde{W}_{t_0+1}) = \arg\max_{\widetilde{c}_t} \mathbb{E}\left[\int_{t_0}^{t_0+1} u(\widetilde{c}_t, t)\mathrm{d}t + \bar{u}(\widetilde{W}_{t_0+1}, t_0+1)\right]$$

subject to the budget constraint

$$d\widetilde{W}_t = \widetilde{W}_t \left(\mu_Y dt + \sigma_Y dB_t\right) - \widetilde{c}_t dt,$$

where $u(\widetilde{c}_t, \widetilde{W}_t) = e^{-\rho t}\widetilde{c}_t^{1-\gamma}(1-\gamma)^{-1}$ and $\bar{u}(\widetilde{W}_{t_0+1}, t_0+1) = e^{-\rho(t_0+1)}I(t_0+1)\widetilde{W}_{t_0+1}^{1-\gamma}(1-\gamma)^{-1}$ for $\rho \geq 0$ and $\gamma > 0$.

The conjectured solution is of the form

$$J(\widetilde{W}_t, t) = e^{-\rho t} I(t) \frac{\widetilde{W}_t^{1-\gamma}}{1-\gamma},$$

yielding the following Bellman equation

$$\max_{c_t} J_{\widetilde{W}} \left(\widetilde{W}_t \mu_Y - \widetilde{c}_t\right) + J_t + \frac{1}{2} J_{\widetilde{W}\widetilde{W}} \widetilde{W}_t^2 \sigma_Y^2 + e^{-\rho t} \frac{\widetilde{c}_t^{1-\gamma}}{1-\gamma} = 0 \tag{3.A.8}$$

subject to the following boundary condition

$$J(\widetilde{W}_{t_0+1}, t_0+1) = e^{-\rho(t_0+1)} I(t_0+1) \frac{\widetilde{W}_{t_0+1}^{1-\gamma}}{1-\gamma}.$$

Eq. (3.A.8) yields the following first order condition for $\widetilde{c}_t^\star$, i.e.,

$$-J_{\widetilde{W}} + e^{-\rho t}(\widetilde{c}_t^\star)^{-\gamma} = 0 \Leftrightarrow \widetilde{c}_t^\star = I(t)^{-\frac{1}{\gamma}} \widetilde{W}_t. \tag{3.A.9}$$

Plugging the value function's partial derivatives together with Eq. (3.A.9) into Eq. (3.A.8) gives

$$\begin{aligned} &e^{-\rho t} I(t) \widetilde{W}_t^{-\gamma} \left(\widetilde{W} \mu_Y - I(t)^{-\frac{1}{\gamma}} \widetilde{W}_t\right) + e^{-\rho t} \frac{\widetilde{W}_t^{1-\gamma}}{1-\gamma}(I'(t) - \rho I(t)) + \\ &\frac{1}{2}(-\gamma) e^{-\rho t} I(t) \widetilde{W}_t^{1-\gamma} \sigma_Y^2 + e^{-\rho t} \frac{1}{1-\gamma} \left(I(t)^{-\frac{1}{\gamma}}\right)^{1-\gamma} \widetilde{W}_t^{1-\gamma} = 0, \end{aligned}$$

which can be simplified to the following ODE

$$I(t) \underbrace{\left(\rho - (1-\gamma)\left(\mu_Y - \frac{\sigma_Y^2 \gamma)}{2}\right)\right)}_{=:\xi} - \gamma I(t)^{\frac{\gamma-1}{\gamma}} = I'(t),$$

or equivalently

$$I(t)^{\frac{1-\gamma}{\gamma}} I'(t) = I(t)^{\frac{1}{\gamma}} \xi - \gamma.$$

If we set $Z(t) := I(t)^{\frac{1}{\gamma}}$, i.e., $Z'(t) = \frac{1}{\gamma}I(t)^{\frac{1-\gamma}{\gamma}}I'(t)$, hence we get the following ODE

$$\gamma Z'(t) = Z(t)\xi - \gamma \Leftrightarrow Z'(t) = Z(t)\underbrace{\frac{\xi}{\gamma}}_{=:\phi} - 1 \tag{3.A.10}$$

with boundary condition $Z(t_0+1) = I(t_0+1)^{\frac{1}{\gamma}}$. The corresponding homogenous ODE is

$$Z'(t) = Z(t)\phi \Rightarrow Z(t) = e^{\phi t}C,$$

where $C \in \mathbb{R}$ is some integration constant.

We can solve the non-homogenous equation in Eq. (3.A.10) by applying the variation of constants, i.e.,

$$Z'(t) = \phi e^{\phi t}C(t) + e^{\phi t}C'(t). \tag{3.A.11}$$

Plugging Eq. (3.A.11) into Eq. (3.A.10) yields

$$\phi e^{\phi t}C(t) + e^{\phi t}C'(t) = e^{\phi t}C(t)\phi - 1 \Leftrightarrow C'(t) = -e^{-\phi t},$$

which has the solution $C(t) = \frac{1}{\phi}e^{-\phi t} + \widetilde{C}$, i.e., we get $Z(t) = \frac{1}{\phi} + e^{\phi t}\widetilde{C}$ and recalling the respective boundary condition yields

$$Z(t_0+1) = \frac{1}{\phi} + e^{\phi(t_0+1)}\widetilde{C} = I(t_0+1)^{\frac{1}{\gamma}} \Leftrightarrow \widetilde{C} = \left(I(t_0+1)^{\frac{1}{\gamma}} - \frac{1}{\phi}\right)e^{-\phi(t_0+1)}.$$

Hence, we have found $Z(t) = \frac{1}{\phi} + e^{\phi(t-(t_0+1))}\left(I(t_0+1)^{\frac{1}{\gamma}} - \frac{1}{\phi}\right)$, which by the above definition of $Z(t)$ gives us

$$I(t) = \left(\frac{1}{\phi} + e^{\phi(t-(t_0+1))}\left(I(t_0+1)^{\frac{1}{\gamma}} - \frac{1}{\phi}\right)\right)^{\gamma},$$

and by Eq. (3.A.9) we thus finally get

$$\widetilde{c}_t^{\star} = \frac{\widetilde{W}_t}{\frac{1}{\phi} + e^{-\phi(t_0+1-t)}\left(I(t_0+1)^{\frac{1}{\gamma}} - \frac{1}{\phi}\right)},$$

where

$$\phi = \frac{1}{\gamma}\left(\rho - (1-\gamma)\left(\mu_Y - \frac{\gamma\sigma_Y^2}{2}\right)\right).$$

This completes the proof. □

Proof of Proposition 2. We first note that due to the definition of $\tilde{c}_t$ in Eq. (3.20), W_{t_0+1}

does not depend on μ_D since additional borrowings are immediately consumed by the representative agent.

For $u(\cdot)$ in Eq. (3.23), we get

$$\frac{\partial}{\partial \mu_D} u(c_s^\star, s) = u(c_s^\star + \bar{c}, s) = e^{-\rho s} \frac{2D_s}{(c_s^\star + \bar{c})^\gamma} > 0. \tag{3.A.12}$$

The discounted expected default costs in Eq. (3.20) equal

$$\mathbb{E}\left[e^{-\rho}\Psi \mathbf{1}_{\left\{\frac{D_{t_0+1}-\Delta W}{Y_{t_0+1}} > \frac{D_{t_0}}{Y_{t_0}}\right\}}\right] = e^{-\rho}\Psi \mathbb{P}\left(\frac{D_{t_0+1}-\Delta W}{Y_{t_0+1}} > \frac{D_{t_0}}{Y_{t_0}}\right), \tag{3.A.13}$$

where $\Delta W := W_{t_0+1} - W_{t_0}$. Rewriting the probability of default and taking the first partial derivative with respect to μ_D yields

$$\begin{aligned} \frac{\partial}{\partial \mu_D}\mathbb{P}\left(\frac{D_{t_0+1}-\Delta W}{Y_{t_0+1}} > \frac{D_{t_0}}{Y_{t_0}}\right) &= \frac{\partial}{\partial \mu_D}\mathbb{P}\left(Y_{t_0+1}\frac{Y_{t_0}}{D_{t_0}} + \Delta W < D_{t_0} e^{\mu_D}\right) \\ &= \frac{\partial}{\partial \mu_D} G\left(D_{t_0} e^{\mu_D}\right) \\ &= g\left(D_{t_0} e^{\mu_D}\right) D_{t_0} e^{\mu_D} \geq 0, \end{aligned} \tag{3.A.14}$$

where $G(\cdot)$ denotes the cdf of $Y_{t_0+1}\frac{Y_{t_0}}{D_{t_0}} + \Delta W$, and we have relied on $\mu_D \perp\!\!\!\perp \Delta W$.

Let $\mu_D^\star$ denote a bounded solution to Eq. (3.20). At $\mu_D^\star$ it then has to hold that Eq. (3.A.14) is strictly positive. Thus, for a bounded solution to Eq. (3.20), the strictly positive marginal utility from higher consumption between t_0 and $t_0 + 1$ needs to be offset by the effect of a strictly increasing default probability. In addition, we have to ensure that the derivative of the power utility in Eq. (3.A.12) and the part in Eq. (3.A.14) can maximally interesect twice (at the maximum and (local) minimum). For this to hold, Eq. (3.A.14) needs to converge faster to zero than Eq. (3.A.12) which is true as Eq. (3.A.14) converges exponentially. Hence, there can be only one interior $\mu_D^\star$. This completes the proof. □

Proof of Corollary 1. Analogous to the proof of Proposition 2, we note that due to the definition of $\tilde{c}_t$ in the restatement of Eq. (3.16) under standard debt, i.e., $c_t = \tilde{c}_t - (\mu_D - R)$, and Eq. (3.17) under puttable debt, i.e., $c_t = \tilde{c}_t - (\mu_D - r)$, W_{t_0+1} does not depend on μ_D since additional borrowings are immediately consumed by the representative agent.

For $u(\cdot)$ in Eq. (3.23), we get

$$\frac{\partial}{\partial \mu_D} u(c_s^\star + \bar{c}, s) = e^{-\rho s} \frac{D_s}{(c_s^\star + \bar{c})^\gamma} > 0. \tag{3.A.15}$$

The discounted expected default costs in Eq. (3.20) are given in Eq. (3.A.13) and the first partial derivative with respect to μ_D is positive as shown in Eq. (3.A.14).

Let $\mu_D^\star$ denote a bounded solution to Eq. (3.20). At $\mu_D^\star$ it then has to hold that Eq. (3.A.14) is strictly positive. Thus, for a bounded solution to Eq. (3.20), the strictly positive marginal utility from higher consumption between t_0 and $t_0 + 1$ needs to be offset by the effect of a strictly increasing default probability. In addition, we have to ensure that the derivative of the power utility in Eq. (3.A.15) and the part in Eq. (3.A.14) can maximally interesect twice (at the maximum and (local) minimum). For this to hold, Eq. (3.A.14) needs to converge faster to zero than Eq. (3.A.15) which is true as Eq. (3.A.14) converges exponentially. Hence, there can be only one interior $\mu_D^\star$. This completes the proof. □

Part III

Bibliography

Bibliography

Abeler, J., A. Becker, and A. Falk (2014). Representative evidence on lying costs. *Journal of Public Economics 113*, 96–104.

Abeler, J., D. Nosenzo, and C. Raymond (2016). Preferences for Truth-telling. *Working Paper*.

Akerlof, G. A. (1983). Loyalty Filters. *American Economic Review 73*(1), 54–63.

Akerlof, G. A. (1989). The Economics of Illusions. *Economics & Politics 1*(1), 1–15.

Akerlof, G. A. and R. E. Kranton (2000). Economics and Identity. *Quarterly Journal of Economics 115*(3), 715–753.

Amir, O., D. G. Rand, and Y. K. Gal (2012). Economic Games on the Internet: The Effect of $1 Stakes. *PLoS ONE 7*(2), 1–4.

Andreoni, J. and D. B. Bernheim (2009). Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects. *Econometrica 77*(5), 1607–1636.

Andreoni, J. and A. L. Sanchez (2014). Do Beliefs Justify Actions or Do Actions Justify Beliefs? An Experiment on Stated Beliefs, Revealed Beliefs, and Social-Image Motivation. *NBER Working Paper*.

Arlot, S. and A. Celisse (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys 4*, 40–79.

Babcock, L. and G. Loewenstein (1997). Explaining bargaining impasse: The role of self-serving biases. *The Journal of Economic Perspectives 11*(1), 109–126.

Barr, D., O. Bush, and A. Pienkowski (2014). GDP-linked Bonds and Sovereign Default. In *Life After Debt*, pp. 246–275. Springer.

Bénabou, R. and J. Tirole (2011). Identity, morals, and Taboos: Beliefs as Assets. *Quarterly Journal of Economics 126*(2), 805–855.

Blanchard, O., P. Mauro, and J. Acalin (2016). The Case for Growth-Indexed Bonds in Advanced Economies Today. *Policy Brief Peterson Institute for International Economics.*

Blanco, M., D. Engelmann, A. K. Koch, and H.-T. Normann (2010). Belief elicitation in experiments: is there a hedging problem? *Experimental Economics 13*(4), 412–438.

Blume, A., D. V. DeJong, Y.-G. Kim, and G. B. Sprinkle (1998). Experimental Evidence on the Evolution of Meaning of Messages in Sender-Receiver Games. *American Economic Review 88*(5), 1323–1340.

Bögli, A., J. Sobel, and A. F. Wagner (2017). Model Selection from Experimental Data: Evidence from Individual Lying Behavior. *Working Paper*.

Bohannon, J. (2016). Mechanical Turk upends social sciences. *Science 352*(6291), 1263–1264.

Bolton, G. E. and A. Ockenfels (2000). ERC: A Theory of Equity, Reciprocity, and Competition. *American Economic Review 90*(1), 166–193.

Borensztein, E., P. Mauro, M. Ottaviani, and S. Claessens (2004). The Case for GDP-Indexed Bonds. *Economic Policy 19 (38)*, 165–216.

Borensztein, E. and U. Panizza (2009). The Costs of Sovereign Default. *IMF Economic Review 56*(4), 683–741.

Breig, Z. (2017). Prediction and Model Selection in Experiments. *Working Paper*.

Brunnermeier, M., L. Garicano, P. R. Lane, M. Pagano, R. Reis, T. Santos, D. Thesmar, S. V. Neiuwerburgh, and D. Vayanos (2011). ESBies: A realistic reform of Europe's financial architecture. *VoxEU.org 25. October.*

Bucciol, A., F. Landini, and M. Piovesan (2013). Unethical behavior in the field: Demographic characteristics and beliefs of the cheater. *Journal of Economic Behavior & Organization 93*, 248–257.

Caballero, R. J. (2003). The Future of the IMF. *American Economic Review 93 (2)*, 31–38.

Cai, H. and J. T. Wang (2006). Overcommunication in Strategic Information Transmission Games. *Games and Economic Behavior 56*(1), 7–36.

Calvo, G. (1988). Servicing the Public Debt: The Role of Expectations. *American Economic Review 78 (4)*, 647–661.

Chance, Z., M. I. Norton, F. Gino, and D. Ariely (2011). Temporal view of the costs and benefits of self-deception. *PNAS 108*(3), 15655–15659.

Charness, G. and M. Dufwenberg (2006). Promises and Partnership. *Econometrica 74*(6), 1579–1601.

Chen, D. L., M. Schonger, and C. Wickens (2015). oTree - An Open-Source Platform for Laboratory, Online, and Field Experiments. *Working Paper*.

Chidambaran, N., C. S. Fernando, and P. A. Spindt (2001). Credit Enhancement through Financial Engineering: Freeport McMoRan's Gold-Denominated Depositary Shares. *Journal of Financial Economics 60*(2), 487–528.

Collard, F., M. Habib, and J.-C. Rochet (2015). Sovereign Debt Sustainability in Advanced Economies. *Journal of the European Economic Association 13*, 381–420.

Collard, F., M. Habib, and J.-C. Rochet (2016). The Reluctant Defaulter: A Tale of High Government Debt. *Working Paper*.

Crawford, V. P. and J. Sobel (1982). Strategic Information Transmission. *Econometrica 50*(6), 1431–1451.

De Grauwe, P. and Y. Ji (2012). Mispricing of Sovereign Risk and Multiple Equilibria in the Eurozone. *CEPS Working Paper*.

Di Tella, R., R. Perez-Truglia, A. Babino, and M. Sigman (2015). Conveniently Upset: Avoiding Altruism by Distorting Beliefs about Others' Altruism. *American Economic Review 105*(11), 3416–3442.

Dickhaut, J. W., K. A. McCabe, and A. Mukherji (1995). An experimental study of strategic information transmission. *Economic Theory 6*, 389–403.

Dufwenberg, M. and M. Dufwenberg (2016). Lies in Disguise - A Theoretical Analysis of Cheating. *Working Paper*.

Dufwenberg, M. and U. Gneezy (2000). Measuring Beliefs in an Experimental Lost Wallet Game. *Games and Economic Behavior 30*(2), 163–182.

Eaton, J. and M. Gersovitz (1981). Debt with Potential Repudiation: Theoretical and Empirical Analysis. *Review of Economic Studies 48 (2)*, 289–309.

Egan, M., G. Matvos, and A. Seru (2017). The Market for Financial Adviser Misconduct. *Working Paper*.

Ellingsen, T. and M. Johannesson (2008). Pride and prejudice: The human side of incentive theory. *American Economic Review 98*(3), 990–1008.

Erat, S. and U. Gneezy (2012). White Lies. *Management Science 58*(4), 723–733.

Ericson, K. M. M., J. M. White, D. Laibson, and J. D. Cohen (2015). Money earlier or later? Simple heuristics explain intertemporal choice better than delay discounting does. *Psychological Science 26*(6), 826–833.

Favero, C. and A. Missale (2012). Sovereign Spreads in the Euro Area: Which Prospects for a Eurobond? *Economic Policy 27 (70)*, 231–273.

Fehr, E. and K. M. Schmidt (1999). A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics 114*(3), 817–868.

Fischbacher, U. and F. Heusi (2013). Lies in Disguise: An Experimental Study on Cheating. *Journal of the European Economic Association 11*(3), 525–547.

Freud, S. (1933). *New Introductory Lectures on Psycho Analysis.* W.W. Norton & Company.

Gali, J., J. D. Lopez-Salido, and J. Vallés (2007). Understanding The Effects of Government Spending on Consumption. *Journal of the European Economic Association 5(1)*, 227–270.

Garbarino, E., R. Slonim, and M. C. Villeval (2017). Loss Aversion and Lying Behavior: Theory, Estimation and Empirical Evidence. *Working Paper*.

Gibson, R., C. Tanner, and A. F. Wagner (2013). Preferences for Truthfulness: Heterogeneity among and within Individuals. *American Economic Review 103*(1), 532–548.

Gneezy, U. (2005). Deception: The Role of Consequences. *American Economic Review 95*(1), 384–394.

Gneezy, U., A. Kajackaite, and J. Sobel (2016). Lying Aversion and the Size of the Lie. *Working Paper*.

Gneezy, U., B. Rockenbach, and M. Serra-Garcia (2013). Measuring lying aversion. *Journal of Economic Behavior & Organization 93*, 293–300.

Gneezy, U., S. Saccardo, M. Serra-Garcia, and R. van Veldhuizen (2015). Motivated Self-deception and Unethical Behavior. *Working Paper*.

Hellwig, C. and T. Philippon (2011). Eurobills, not Eurobonds. *VoxEU.org 02 December.*

Horton, J. J., D. G. Rand, and R. J. Zeckhauser (2011). The online laboratory: conducting experiments in a real labor market. *Experimental Economics 14*(3), 399–425.

Hutton, A. P., L. F. Lee, and S. Z. Shu (2012). Do Managers Always Know Better? The Relative Accuracy of Management and Analyst Forecasts. *Journal of Accounting Research 50*(5), 1217–1244.

Kartik, N. (2009). Strategic Communication with Lying Costs. *Review of Economic Studies 76*(4), 1359–1395.

Kerschbamer, R. and M. Sutter (2017). The Economics of Credence Goods - a Survey of Recent Lab and Field Experiments. *CESifo Economic Studies 63*(1), 1–23.

Khalmetski, K. and D. Sliwka (2017). Disguising Lies - Image Concerns and Partial Lying in Cheating Games. *CESifo Working Paper No. 6347*.

Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *International Joint Conference on Artificial Intelligence 14*, 1137–1145.

Köszegi, B. and M. Rabin (2006). A Model of Reference-Dependent Preferences. *Quarterly Journal of Economics 121*(4), 1133–1165.

Lane, P. R. (2012). The European Sovereign Debt Crisis. *Journal of Economic Perspectives 26 (3)*, 49–68.

López-Pérez, R. and E. Spiegelman (2013). Why do people tell the truth? Experimental evidence for pure lie aversion. *Experimental Economics 16*(3), 233–247.

Mazar, N., O. Amir, and D. Ariely (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research 45*, 633–644.

Mele, A. R. (1997). Real self-deception. *Behavioral and Brain Sciences 20*, 91–136.

Merton, R. C. (1969). Lifetime Portfolio Selection under Uncertainty: The Continuous-Time Case. *The Review of Economics and Statistics 51*, 247–257.

Merton, R. C. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *The Journal of Finance 29*(2), 449–470.

Merton, R. C. (1977). An Analytic Derivation of the Cost of Deposit Insurance and Loan Guarantees. *Journal of Banking and Finance 1*, 3–11.

Müller, A., K. Storesletten, and F. Zilibotti (2016). Sovereign Debt and Structural Reforms. *Working Paper*.

Neftci, S. N. and A. O. Santos (2003). Puttable and Extendible Bonds: Developing Interest Rate Derivatives for Emerging Markets. *IMF Working Paper*.

Paolacci, G., J. Chandler, and P. G. Ipeirotis (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making 5*(5), 411–419.

Peysakhovich, A. and J. Naecker (2017). Using Methods from Machine Learning to Evaluate Behavioral Models of Choice Under Risk and Ambiguity. *Journal of Economic Behavior and Organization 133*, 373–384.

Plato (1953). *Cratylus. In: The dialogues of Plato, trans. B. Jowett.* Clarendon Press.

Rabin, M. (1995). Moral Preferences, Moral Constraints, and Self-Serving Biases. *Working Paper*.

Rochet, J.-C. (2006). Why do Countries Default? *Working Paper*.

Rode, J. (2010). Truth and trust in communication: Experiments on the effect of a competitive context. *Games and Economic Behavior 68*(1), 325–338.

Sánchez-Pagés, S. and M. Vorsatz (2007). An Experimental Study of Truth-Telling in a Sender-Receiver Game. *Games and Economic Behavior 61*(1), 86–112.

Schwardmann, P. and J. van der Weele (2016). Deception and Self-Deception. *Working Paper*.

Schweitzer, M. E. and C. K. Hsee (2002). Stretching the Truth: Elastic Justification and Motivated Communication of Uncertain Information. *Journal of Risk and Uncertainty 25*(2), 185–201.

Shalvi, S., F. Gino, R. Barkan, and S. Ayal (2015). Self-serving Justifications: Doing Wrong and Feeling Moral. *Current Directions in Psychological Science 24*(2), 125–130.

Shiller, R. J. (1994). *Macro Markets: Creating Institutions for Managing Society's Largest Economic Risks.* Oxford University Press.

Shiller, R. J. (2003). *The New Financial Order: Risk in the 21st Century.* Princeton University Press.

Sutter, M. (2009). Deception through Telling the Truth?! Experimental Evidence from Individuals and Teams. *Economic Journal 119*(534), 47–60.

Tanner, C., B. Ryf, and M. Hanselmann (2009). Geschützte Werte Skala: Konstruktion und erste Validierung eines Messinstrumentes (Protected Values Measure: Construction and First Validation of an Instrument to Assess Protected Values). *Diagnostica 55*(3), 174–183.

Van Lange, P. A. M., W. Otten, E. M. N. D. Bruin, and J. A. Joireman (1997). Development of Prosocial, IndIndividual, and Competitive Orientations: Theory and Preliminary Evidence. *Journal of Personality and Social 73*(4), 733–746.

Vanberg, C. (2008). Why Do People Keep Their Promises? An Experimental Test of Two Explanations. *Econometrica 76*(6), 1467–1480.

Wang, J. T., M. Spezio, and C. F. Camerer (2010). Pinocchio's Pupil: Using Eyetracking and Pupil Dilation to Understand Truth Telling and Deception in Sender-Receiver Games. *American Economic Review 100*(3), 984–1007.

Weibull, J. and E. Villa (2005). Crime, punishment and social norms. *SSE/EFI Discussion Paper*.

Part IV

Curriculum Vitae

Curriculum Vitae

Personal details

Name:	Andrin T. Bögli
Date of Birth:	January 19, 1989
Nationality:	Swiss

Education

09/2011 – 10/2017	PhD Program in Finance at the Swiss Finance Institute University of Zurich (Zurich, Switzerland) Supervisor: Prof. Dr. Alexander F. Wagner
07/2016 – 12/2016	Visiting PhD Student Harvard University (Cambridge MA, USA)
04/2016 – 07/2016	Visiting PhD Student University of California, San Diego (La Jolla CA, USA)
09/2010 – 04/2016	MSc Banking and Finance University of Zurich (Zurich, Switzerland)
09/2007 – 10/2010	BA Banking and Finance University of Zurich (Zurich, Switzerland)

Professional Experience

09/2012 – 10/2017	Research and Teaching Assistant at the University of Zurich (Zurich, Switzerland)
01/2010 – 08/2010	Internship at UBS Investment Bank (Opfikon, Switzerland)